Language in the hands

Mol, L.

Publication date:
2011

Document Version
Publisher's PDF, also known as Version of record

Language in the hands

Lisette Mol

Language in the hands

Lisette Mol
PhD thesis
Tilburg University, 2011

TiCC Ph.D series no. 18

ISBN/ EAN: 978-90-8570-429-4
Print: CPI Wöhrmann print service
Cover design: Hans Westerbeek

Language in the hands

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan Tilburg University
op gezag van de rector magnificus,
prof. dr. Ph. Eijlander,
in het openbaar te verdedigen ten overstaan van
een door het college voor promoties aangewezen commissie
in de aula van de Universiteit
op maandag 7 november 2011 om 16:15 uur
door

Elisabeth Margaretha Maria Mol

geboren op 2 juni 1981 te Hoorn

Promotores:

Prof. dr. E.J.Krahmer
Prof. dr. A.A. Maes
Prof. dr. M.G.J. Swerts

Promotie-commissie:

Prof. Dr. S.E. Brennan
Prof. Dr. J.P. de Ruiter
Dr. S. Kita
Prof. Dr. A. Özyürek
Dr. W.M.E. van de Sandt – Koenderman

Contents

Chapter 1

Introduction

In a recent movie by Quentin Tarantino, an American man pretends to be German. As I watched the movie in a theater, I caught my breath when he asked for three glasses with a bottle of whiskey. Not that anything was wrong with his pronunciation of *drei Gläser* (three glasses). Yet he held up three fingers while ordering: his index, middle, and ring finger! This easily gave away his disguise, since in Germany the common way of gesturing *three* is by extending the thumb, index, and middle finger. Although I was worried about the character being exposed, I was at the same time quite pleased that gesture played such an important part on the big screen.

The fact that our hands are somehow involved in communication seems to be universal. Until today, not a single language or culture is known in which people do not gesture. While speaking, most people tend to move their hands around, seemingly indicating references, simulating action or movement, depicting objects or concepts, placing emphasis… Though very common, most of these movements seem less conventional than speech is (Kendon, 2004). What role do these speech accompanying hand gestures play? Do they somehow aid speech production, do they regulate our interactions, or might they speak for themselves?

The idea that our hands can (almost) be said to speak goes back at least as far as the Roman era, when it was mentioned in the *Institutio Oratoria* ('Education of the Orator'), written by Marcus Fabius Quintilianus in the first century AD (as described in Kendon, 2004). Quintilianus recognized that hands could be used to communicate what we want from others (such as by demanding or pleading), express attitudes or feelings (e.g. joy, sorrow, hesitation, approval) and indicate concrete things such as measurement, quantity, number and time. Clearly, Quintilianus assumed a communicative purpose of gesture. Yet although he recognized that hands can sometimes communicate on their own, without speech, his work mostly describes how to best use the hands and body so as to comment on and help convey ideas that are predominantly expressed in speech.

The idea that gesture is somehow subordinate to speech is still a popular view today, albeit for different reasons. According to some studies, our hands rather literally serve speech (e.g. Krauss, 1998; Krauss, Chen, & Gottesman, 2000). That is, producing gestures may aid the retrieval of lexical or phonological forms, which eventually leads to the articulation of speech. Producing hand gestures may also play a role in the process of thinking for speaking (De Ruiter, 1998; Kita, 2000), or support cognition more generally (Chu & Kita, 2008;

Goldin-Meadow, 2010; Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001). What these views have in common is that gesture production first and foremost serves the person producing the gesture.

This is not to say that addressees cannot benefit from seeing gestures as well. Yet some theories go beyond this accidental benefit and propose that gesture is meant to convey information. For example, gestures may play a role in regulating interaction (Bavelas, Chovil, Lawrie, & Wade, 1992). Speakers may also gesture to convey part of their message. Kendon (2004) regards both speech and gesture as part of a speaker's communicative effort. McNeill (2005) even assumes a single process underlying both gesture and speech production. In his growth point theory, an idea arises (the growth point) and is then developed as it is translated into an utterance, which will be expressed in speech and gesture jointly. In these latter two views, gesture is on par with speech in it being communicative and it being intended as such.

This dissertation addresses the question of whether the role of gesture is limited to facilitating speech, or whether it goes beyond that, with gesture (like speech) being part of language production itself. The first two studies assess whether speakers have a communicative intent with their gesturing. We test whether speakers adapt their gesturing to their beliefs about their addressee. For example, if speakers gesture to communicate, then they will gesture differently depending on whether they believe their addressee can see them or not, independent of any changes in the speaker's environment. The same may hold for whether a speaker believes to be addressing another person, or an artificial system. If speakers gesture in the same way under all these circumstances, then it is unlikely that gesture is intended communicatively.

The final two studies examine whether gesture behaves like speech in two more ways. In our third study, we investigate what happens when interlocutors spontaneously copy each other's gesture forms. We compare this copying of gestures to the spontaneous copying of other nonverbal behaviors as well as to the repetition of each other's words. If gesture acts like speech, the repetition of gesture forms across interlocutors will resemble the repetition of words. In our fourth and final study, we test what happens to gesture when speech is impaired as a result of stroke, specifically, we examine the gestures of aphasic speakers. Can gesture be used as an alternative means of communication by speakers with aphasia? Or is it too affected by this language disorder?

Contrary to the earliest studies on gesture, our goal is not to prescribe how gesture is best adapted to our thought or to our narrative, but rather to look at how gestures are commonly produced by speakers. The methods employed are empirical in nature. By examining how speakers gesture under different circumstances, how speakers adapt their gesturing to one another, and how gesture is affected by impairment of speech due to brain damage, we aim to inform theories on why speakers produce gestures, how gesture relates to thought, and how gesture and speech production are interrelated. Clearly, these questions are too big to be answered by just the four studies described in this dissertation. Yet each in their own way, the studies contribute to our understanding of gesture production.

Dissertation outline

This dissertation contains four studies, which have been published, have been accepted for publication, or have been submitted for publication as full papers in scientific journals. Being self-contained, each of the next four chapters has its own abstract, introduction, discussion and reference list. Chapter 6 contains a general discussion and conclusion.

References

Bavelas, J., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse Processes, 15*, 469-489.

Chu, M., & Kita, S. (2008). Spontaneous gestures during mental rotation tasks: Insights into the microdevelopment of the motor strategy. *Journal of Experimental Psychology: General, 137*(4), 706-723.

De Ruiter, J. P. (1998). Gesture and Speech Production. Unpublished Doctoral Dissertation. University of Nijmegen.

Goldin-Meadow, S. (2010). When gesture does and does not promote learning. *Language and Cognition, 2*(1), 1-19.

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining Math: Gesturing Lightens the Load. *Psychological Science, 12*(6), 516-522.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and Gesture*. Cambridge: Cambridge University Press.

Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science, 7*, 54-60.

Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical acces: A process model. In D. McNeill (Ed.), *Language and gesture* (pp. 261-283). New York: Cambridge University Press.

McNeill, D. (2005). *Gesture and Thought*. Chicago and London: University of Chicago Press.

Chapter 2

The communicative import of gestures: evidence from a comparative analysis of human-human and human-machine interactions

Abstract
Does gesturing primarily serve speaker internal purposes, or does it mostly facilitate communication, for example by conveying semantic content, or easing social interaction? To address this question, we asked native speakers of Dutch to retell an animated cartoon to a presumed audiovisual summarizer, a presumed addressee in another room (through web cam), or an addressee in the same room, who could either see them and be seen by them or not. We found that participants gestured least frequently when talking to the presumed summarizer. In addition, they produced a smaller proportion of large gestures and almost no pointing gestures in this setting. Two perception experiments revealed that observers are sensitive to this difference in gesturing. We conclude that gesture production is not a fully automated speech facilitation process, and that it can convey information about the communicative setting a speaker is in.

This chapter is based on:
Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2009). The communicative import of gestures: Evidence from a comparative analysis of human-human and human-computer interactions. *Gesture*, *9*(1), 97-126.

Introduction

In this chapter, we explore the functional roles of spontaneous hand gestures produced during narrative speech, by looking at it from the production as well as the perception perspective. If gesture production primarily serves speaker internal processes, then we would expect it to be a highly automated process that is little influenced by the communicative setting a speaker is in, and by whom a speaker is addressing. On the other hand, if gestures primarily aid communication between a speaker and an addressee, then we would expect gesturing to be a more flexible process, which is adapted to different communicative environments and audience characteristics.

If speakers gesture mostly for themselves, then addressees may or may not be able to use information from gestures. However, if addressees are unable to use information from the gesture modality, this would make it less likely that speakers intend their gestures communicatively. Since people continuously switch roles between speaker and addressee in day-to-day communication, we think it unlikely that speakers would put communicative effort into a modality that they never use in comprehension. And for the same reason, if speakers gesture partly to communicate, then we would expect that addressees are able to gain information from speakers' gestures.

In this chapter we describe two studies. First we describe an experiment from the speaker's perspective, in which we manipulated the nature of the addressee (either artificial or human) and whether the speaker and addressee could see each other. We were interested in the effects of these manipulations on gesture production. Our second study consists of two perception experiments using video-clips from the first study. We measured whether addressees were sensitive to possible differences in gesturing that resulted from a speaker addressing a human or an artificial addressee.

Background

The functional role of gestures

Many studies have been conducted to investigate the primary functional role of hand gestures. One view is that gestures are mostly produced for the benefit of the speaker, for example to aid speech production. Many studies have found

evidence in support of this view (e.g. De Ruiter, 1998; Hadar, 1989; Hostetter, Alibali, & Kita, 2007; Hostetter & Hopkins, 2002; Kita, 2000; Krauss, 1998). Other studies have shown that gesturing may facilitate cognition in processes other than language production, which is another for-speaker function (Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001; Goldin-Meadow & Sandhofer, 1999). Another view is that speakers produce gestures with a communicative intent. Kendon (2004) for example, argues that speakers produce gestures as an integral part of their communicative effort. Much support for this hypothesis has been found (Bangerter & Chevalley, 2007; Cohen, 1977; Cohen & Harison, 1973; Jacobs & Garnham, 2007; Özyürek, 2002), also see the review paper by Kendon (1994).

Alibali, Heath, and Myers (2001) have tried to reconcile various seemingly contradictory experimental results by associating different types of gestures with different functional roles. They conducted a study in which narrators told a story to an addressee either face-to-face, or with an opaque screen in between speaker and addressee. They found that speakers produced more representational gestures (gestures that depict some of the content of the story) in the face-to-face condition than in the screen condition, when the addressee could not see the speaker, although representational gestures were also produced in this condition. Beat gestures (biphasic gestures that do not depict narrative content) on the other hand, were produced at comparable rates under both conditions. The fact that speakers still produced many (representational) gestures when it was clear that the addressee could not see them is not easily explained by a theory that stresses the communicative function of gestures. Alibali et al. concluded that both types of gesture serve both speaker-internal and communicative functions. They suggested examining 'how different speakers use gestures in different types of contexts for both speaker-internal and communicative purposes' rather than trying to find a single primary role of gesture production. We will briefly review several factors that have been suggested to affect gesture production, which are relevant to the present study.

Factors influencing gesture production

Visibility and dialogue Bangerter and Chevalley (2007) investigated the effect of mutual visibility on pointing gestures in a referential communication task. They found that pointing movements that did not involve raising the arm, were

produced at equal rates, regardless of whether conversational partners could see each other or not. This suggests that they are automatic in production. However, pointing movements that did involve raising the arm were used more when interlocutors could see each other, suggesting that they are intended to communicate. Thus, gesture size seems to be indicative of the gesture's functional role, and of the nature of the cognitive processes underlying its production.

In a somewhat similar vein, Enfield, Kita, and De Ruiter (2007) describe a theory of how different sizes of pointing gestures serve different pragmatic functions in face-to-face communication. Based on data from the language Lao, they argue that larger pointing gestures carry primary, "informationally foregrounded" information, whereas smaller pointing gestures carry "informationally backgrounded information, which refers to a possible but uncertain lack of referential common ground".

The importance of gesture size in relation to visibility was also found by Bavelas, Gerwing, Sutton, and Prevost (2008). In a picture description task, they compared face-to-face communication (which enables dialogue and visibility) to talking through a hand held phone (dialogue, but no visibility) and talking to a tape recorder using a hand held microphone (no dialogue, no visibility). They found that speakers gestured more while being engaged in dialogue, and also that they gestured very differently if there was the possibility to demonstrate things to the addressee by gesture. Participants described a picture of an old-fashioned dress. In the face-to-face condition, gestures were done to describe features of the dress as if it was full size. In the phone condition, gestures were only the size of the picture, and proved harder to interpret. In the tape recorder condition, gestures were very small and it was hard for the coders to interpret their meaning. Thus, visibility had a large effect on how people gestured and the presence of dialogue had a large effect on gesture rate.

Listener needs Besides mutual visibility and dialogue, Jacobs and Garnham (2007) point out that gesture production may depend on the behavior and needs of the addressee (also see Enfield, et al., 2007) and on the type of task that the speaker is performing. They found that narrators produced fewer gestures when they knew that their addressee already knew part of the content of the story they were telling. They also found that speakers produced more gestures when the addressee appeared attentive, than when the addressee appeared inattentive. They

therefore concluded that during narrative tasks, gestures are produced primarily for the benefit of the addressee.

Content Melinger and Levelt (2004) looked at the type of information being represented. They found that speakers who used gestures representing spatial information omitted more critical spatial information from their verbal descriptions than speakers who did not gesture. They showed that some speakers divided information between the gesture and speech modality. This shows that co-speech gestures expressing spatial information can be used communicatively.

Hostetter and Hopkins (2002) have shown that speakers accompanied their narration with more representational gestures (which they term "lexical movements") if they watched an animated cartoon and subsequently were asked "to picture the events they saw in the cartoon in their head and then describe them" (p. 25), than when they read a description of the events in the cartoon and were asked "to picture the words as they had read them on the page and then relate them" (p. 25) while retelling the events. They interpret this as evidence that representational gestures (lexical movements) are produced more frequently when expressing a thought that is encoded spatially, than when expressing a thought that is encoded textually.

Human-machine interaction

Next to the above-described factors that influence gesture production, human-machine interaction is an important factor in our present study as well. Reeves and Nass (1996) state that "people's responses to media are fundamentally social and natural". This is the so-called media equation and it applies to everyone. They state that the confusion of mediated life and real life is not rare and inconsequential, and that it cannot be corrected with age, education, or thought. Even though their studies focused on social responses, e.g. empathy, rather than on communicative behavior, this would suggest that, even if gestures are used to communicate, people would still gesture at computers and other media, since their social responses may underlie their communication.

Although people show social responses to media and artificial agents, one can ask whether they do so to the same extent as to human interlocutors, and how exactly this influences their communicative behavior. Aharoni and Fridlund (2007) conducted a study in which participants smiled more and used more silence fillers to a purported human interviewer than to a computer interviewer.

In both cases a prerecorded stimulus was used. They found that simply labeling the stimulus as 'human' caused people to be more communicative. In addition, Maes, Marcelis and Verheyen (2007) showed that if speakers assume that their addressee is human, more referential effort will be made than if they assume the addressee is a computer. Respondents more frequently described more attributes than necessary to identify an object to the presumed human addressee, than towards the presumed computer. These findings suggest that at least in some cases, people are wordier toward human than toward computer addressees.

Present study

We are interested in the effect of the addressee being human or artificial on gesture production, and in whether possible differences in gesturing resulting from this manipulation are informative to naïve observers. This is because the different functional roles that gesture may serve imply different predictions on how people would gesture toward an artificial addressee, and place different requirements on addressees' sensitivity to differences in gesturing. We will first describe our production study and then our perception study.

If gesturing is mostly a for-speaker process, either facilitating language production or supporting cognition in another way, then with a similar task, we would expect speakers to gesture in the same way, regardless of the addressee. On the other hand, if gestures are produced to communicate or if gesturing is tied to other aspects of human dialogue, then the addressee being human or artificial may very well influence gesture production. Therefore, we compared a condition in which there was a human addressee with a condition in which there was an artificial addressee, keeping other factors as similar as possible.

For this we have made use of computer mediation. We created a situation similar to one-way video conferencing. A speaker was filmed and was told that an addressee was watching the recordings live, in another room. Throughout this chapter we refer to this condition as the 'Web cam condition'. In this condition there was one-way visibility, no physical co-presence, no dialogue, but the speaker believed there was a human addressee. In a very similar condition, the speaker was told instead that the audiovisual signal of the camera went to an audiovisual summarizer that was located in another room. This condition has similar one-way 'visibility' and, as in the Web cam condition, there was no physical co-presence and no dialogue. Yet this time, the speaker believed there was an artificial addressee.

In both of these settings, participants were asked whether they understood whom they were addressing, before they started their narration. Only if this was clear to them did the experiment proceed. This is different from the tape-recorder condition in the experiment by Bavelas et al. (2008), in which participants were excluded if they had imagined an addressee. Thus, in their tape-recorder condition the addressee was absent entirely rather than artificial. With this design we have also been able to separate the effects of being visible to an (artificial) addressee from the effect of dialogue, since we have been able to introduce a condition in which the speaker could be seen by another person, yet there was no possibility of dialogue.

To control for the effects of physical co-presence and mutual visibility, which are absent in both the condition with the artificial addressee and the condition with a human addressee in another room, we have included two more conditions in our design. These were the conditions used in Alibali et al. (2001): face-to-face communication, in which there is a human addressee, physical co-presence and mutual visibility, and a condition in which speaker and addressee are in the same room, but separated by an opaque screen. Although both of these conditions enable dialogue, we prevented true dialogue from happening by instructing addressees not to interrupt the speaker, but to act naturally otherwise. Thus, addressees were looking at the speaker and gave occasional non-verbal feedback, but they tried to avoid speaking themselves.

For our production study, we asked speakers to retell an animated cartoon in which there are many actions involving direction and moving from one location to another. According to Hostetter and Hopkins (2002), this should lead speakers to produce many representational gestures. And based on the results found by Melinger and Levelt (2004), we would expect speakers to use gestures that express spatial information communicatively in this narration task. Content was always said to be new to the addressee and, as explained above, addressees were instructed not to interrupt the speaker. This was in order to minimize the effects found by Jacobs and Garnham (2007).

Based on the results by Aharoni and Fridlund (2007) as well as the results found by Maes et al. (2007), and the assumption that gesturing bears a communicative function, we would expect participants to produce more gestures in our conditions with a human addressee, than in our condition with an artificial addressee. In addition, based on previous results with imagistic gestures (Bavelas, et al., 2008) and pointing gestures (Bangerter & Chevalley, 2007) we

would expect representational gestures to be larger in conditions in which they have communicative potential.

As mentioned in the introduction, we think it unlikely that speakers would put communicative effort into the gesture modality if they never use this modality as a source of information. In addition, a possible difference between gesturing to a human or to an artificial addressee cannot play a significant role in interaction if addressees are ignorant to this difference. We therefore also conducted a perception study, in which we asked participants to judge whether a speaker was talking to a human or to an artificial addressee, based on movie clips from our production study. These clips were played without sound and different conditions included or excluded the hands and face of the speaker.

Production study

Method

Design As outlined in the previous section under 'Present study', we have used a between subjects design with four conditions. A schematic overview of the settings can be found in Figure 1. We are mainly interested in the effect of the addressee being human or artificial, which is the only difference between our Computer condition and Web cam condition. In both conditions the speaker does not receive any feedback from the addressee.

So that we may see how closely communication through web cam resembles face-to-face communication, we have added a Screen and Face-to-Face condition. Addressees in these conditions were instructed not to interrupt the speaker, but to act naturally otherwise, in order to enhance similarity with the aforementioned conditions. In contrast to the Web cam condition, in both of these conditions the addressee was seated in the same room as the speaker. In the Face- to-Face condition there is mutual visibility as well. If either physical co-presence or seeing the addressee plays a critical role in performing the narration task, then this should result addressee plays a critical role in performing the narration task, then this should result in notable differences between the Web cam, Face-to-Face, and Screen condition. By comparing these three conditions, we get an idea of how closely the Web cam condition resembles a condition with a true and physically present human addressee.

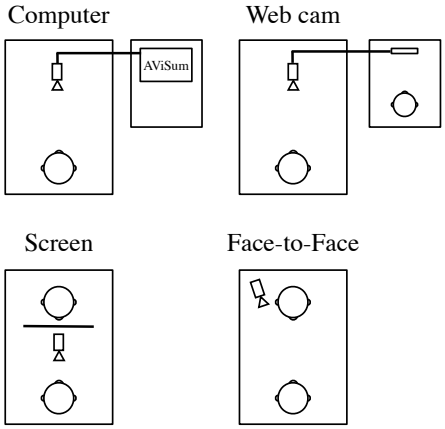

Figure 1: Experimental settings.

Participants Forty-three participants volunteered as narrators for this study. We excluded three participants, because they either were suspicious about the experimental setup or ignored the instructions. The remaining 40 participants (10 male, 30 female) were between the age of 17 and 48 (M = 23, median 19). They were all native speakers of Dutch. None of the participants objected to being recorded, and all of them consented to their data being used for research and educational purposes. There were 11 participants in the Computer condition, 10 in the Web cam condition, 9 in the Screen condition, and 10 in the Face-to-Face (FtF) condition. The listeners in the Screen and FtF condition were confederates.

Procedure We randomly assigned participants to one of four conditions. Narrators first read the instructions (see below for more detail) and could ask any questions they had on the task. The instructions focused on the task of the addressee, namely to summarize the speaker's narration. This way we suggested that the study was on summarizing. Speakers were explicitly asked not to summarize themselves, but to just retell the story. They then watched a seven minute animated cartoon called "Canary Row", which we chose because it has proven to elicit gestures in several other studies, such as McNeill (1992) and Alibali et al. (2001). After being seated in front of the camera, in the Computer and Web cam condition the experimenter asked whether the participant had understood whom they were going to talk to, and paraphrased their answer if it was correct and elaborated on it if it was incomplete. In the Screen and Face-to-

Face condition, the experimenter repeated that the speaker was not to address the camera, but the other participant.

In the Computer condition, the written instructions stated that the signal of the camera was sent to a beta version of an audiovisual summarizer (AViSum) that was located in another building on campus, and which would produce a summary of their narration afterward. It was emphasized that the system could process both auditory and visual information. A fake phone call was made by the experimenter to check whether the signal was received well, and whether the system was ready for use. In reality, there was no such computer system. However, it is not inconceivable that such a system could exist. Dupont & Luettin (2000), for example, describe a speech recognition system that uses both acoustic and visual speech information, and McCowan et al. (2005) describe how automatic analysis of meetings can benefit from information from the visual modality.

In the Web cam condition, the instructions said that the camera was used as a web cam, and that another participant was watching the speaker in another campus building, with the purpose of summarizing their narration afterwards. The experimenter pretended to set up a one-way videoconference with a presumed experimenter in the other building, and then made a fake phone call to check whether the image and sound were received well and whether they were ready to begin. In reality, there was no other participant watching.

In the Screen condition, two students came to the lab, one of which was a confederate. The experimenter pretended to randomly assign the roles of speaker and listener, but always assigned the true participant the role of speaker. After the participant had watched the animated cartoon, narrator and addressee were allowed to ask any questions they had about the task. A wooden screen separated them, such that they could not see each other during the story telling. The narrators' instructions stated that the addressee had to summarize the story afterward, and that they were videotaped with the purpose of comparing the addressee's summary to their narration. We instructed addressees not to interrupt the narrator, but to act naturally otherwise. Occasionally, there was some auditory feedback (laughs, occasional uh-huh's). The Face-to-Face condition differed from the Screen condition only in that narrators retold the story in a face-to-face situation, without the screen in between narrator and addressee.

In each condition, participants were videotaped using a digital video camera. They were seated in front of the camera. The camera position was such that the

entire upper part of the body was visible, including the upper legs. In all conditions, the narrator could look at snapshots of each of the episodes of the cartoon that hung either on the wall or on the screen in front of them. This was done in order to aid memory, and to facilitate more structured, and more comparable stories.

After retelling the cartoon, in the Screen and Face-to-Face condition the experimenter first took the addressee to another room, supposedly to write the summary. Narrators then completed a questionnaire, which included questions on how they had experienced the conversation and whether they had believed the experimental setup. We fully debriefed all participants and asked their consent to use the recordings. The experimenter also asked whether the participants had believed the experimental setup and whether they had suspected any deception.

Transcribing and coding The first author transcribed each narration from the videotape. Repairs, repeated words, false starts, and filled pauses were included. The annotation of gestures was done blind to condition and initially by the first author. Difficult cases were resolved by discussion among the authors.

Initially, coding concentrated on movements of the hands. Later on, when coding for gesture size, movements of other body parts were considered, but only if they occurred simultaneously with a hand gesture. We first discriminated between gestures and other movements such as self-adjustment. We then coded gestures according to McNeill (1992, pp. 78-82), but adding interactive gestures as a category (Bavelas, Chovil, Lawrie, & Wade, 1992). Gestures were first coded as representational, beat, or interactive. This first division could largely be made based on the shape of the gesture. Simple, biphasic movements of the hands were labeled as beat rather than interactive (in Bavelas's definition, interactive gestures subsume the category of beats). Subsequently, we further divided representational gestures into imagistic (iconic or metaphoric) and pointing gestures. Our most important criterion for labeling a gesture as a pointing gesture was the shape of the hand, which should have one or more fingers extended as an index. In addition, we have judged for each of those gestures whether it seemed to only express information on location or direction, or whether it additionally seemed to express information about manner or path. If the latter was the case, the gesture was counted both as imagistic and pointing gesture. Thus, all representational gestures that were not just pointing gestures were counted as imagistic gestures.

In a separate round of gesture coding, we coded for gesture size. Gestures that were produced using only the fingers received a score of 1. If the wrist was moved significantly the gesture received a score of 2. Gestures that also involved significant movement of the elbow or lower arm received a score of 3, and gestures in which the upper arm was also used in a meaningful way, or that involved movement of the shoulder received a score of 4.

Statistical analysis For all tests for significance we used univariate analysis of variance (ANOVA), with condition as the fixed factor (levels: computer, web cam, screen and face-to-face) and a significance threshold of .05. For pairwise comparisons we used the least significance difference test (Fisher, 1951).

Results

Gesture rate Condition had a significant effect on the number of gestures produced per 100 words, $F(3, 36) = 6.27$, $p < .01$, $\eta_p^2 = .34$, see Figure 2. Pairwise comparisons showed that gestures were significantly less frequent in the Computer condition ($M = .64$, $SD = .84$) than in the Web cam ($M = 3.84$, $SD = 4.30$), Screen ($M = 3.69$, $SD = 1.89$), and Face-to-Face (FtF) condition ($M = 6.40$, $SD = 3.86$). The differences between the mean gesture rates in the Web cam, Screen, and FtF conditions were not significant.

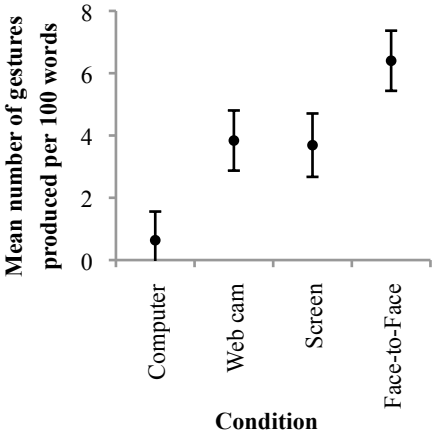

Figure 2: Mean number of gestures produced per 100 words in each condition. Bars represent standard errors.

When performing the analysis with gestures per second rather than per word, we found that gestures were reliably more frequent in the FtF condition (M = .22, SD = .13), than in the Screen (M = .12, SD = .06) and Web cam condition (M = .12, SD = .14), $F(3, 36) = 7.04$, $p < .001$, $\eta_p^2 = .37$. In this analysis too, significantly fewer gestures were produced in the Computer condition (M =.02, SD = .02) than in any of the other three conditions.

Four of the eleven participants in the Computer condition did not produce any gestures. In the other conditions there were no participants that did not gesture at all.

Gesture rate and type We also found a significant effect of condition on representational gestures per 100 words, $F(3, 36) = 5.66$, $p < .01$, $\eta_p^2 = .32$, see Figure 3. Representational gestures were produced at a reliably lower rate in the Computer condition (M = .37, SD = .55) than in the Web cam (M = 2.88, SD = 3.48) and FtF condition (M = 4.79, SD = 3.20). There was a trend toward significance for the difference between the Computer and Screen condition (M = 2.38, SD = 1.37), p = .08. In the Screen condition, reliably fewer representational gestures were produced than in the FtF condition.

For non-representational gestures per 100 words, we found a significant effect of condition as well, $F(3, 36) = 4.75$, $p < 0.01$, $\eta_p^2 = .28$, see Figure 4.

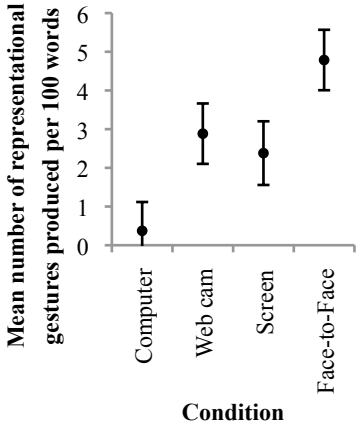

Figure 3: Mean number of representational gestures produced per 100 words in each condition. Bars represent standard errors.

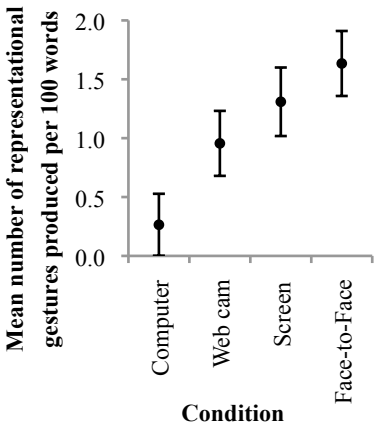

Figure 4: Mean number of non-representational gestures produced per 100 words in each condition. Bars represent standard errors.

Non-representational gestures were produced at a significantly lower rate in the Computer condition (M = .26, SD = .43) than in the Screen (M = 1.31, SD = .78), and FtF condition (M = 1.63, SD = 1.22). There was a trend toward significance for the difference between the Computer and Web cam condition (M = .96, SD = .90), p = .08. In all conditions, representational gestures occurred more frequently than non-representational gestures.

For imagistic gestures, condition had a significant effect on the mean gesture rate, $F(3, 36) = 5.01$, $p < .01$, $\eta_p^2 = .29$. In the Computer condition (M = .37, SD = .55), imagistic gestures were produced significantly less frequently than in the Web cam (M = 2.46, SD = 2.97) and FtF condition (M = 4.01, SD = 2.88). There was a trend toward significance for the difference between the Screen (M = 2.07, SD = 1.16) and FtF condition, p = .06.

Only one pointing gesture was produced in the Computer condition. This was a combined imagistic/ pointing gesture. There was an effect of condition on the number of pointing gestures per 100 words, $F(3, 36) = 4.82$, $p < .01$, $\eta_p^2 = .29$. Pairwise comparisons showed that there was a significant difference between the Screen (M = .50, SD = .65) and FtF condition (M = 1.24, SD = .80). There was a trend toward significance for the difference between the FtF and Web cam condition (M = .61, SD = 1.06), p = .06. The Computer condition (M = .03, SD = .10) differed significantly from the FtF condition and there was a trend toward significance for the difference between the Computer and Web cam condition, p

= .08. When combined imagistic/ pointing gestures were excluded from the analysis, we found similar results. Figure 5 shows the mean number of gestures per 100 words for the different gesture types. In the first bar for pointing gestures, pointing gestures that also seemed to convey significant information on manner or path (imagistic/ pointing gestures) are included, in the second they are not.

Gesture size Using the coding system described in previously, we computed a score that represented the average size of a gesture for each participant. For each participant, we took the sum of the scores of all gestures and divided this sum by the number of gestures produced by that participant. Although overall gesture size did not differ significantly across conditions, $F(3, 32) = 1.34$, $p = .28$, there was a tendency for gestures to be larger in the conditions where speakers thought that the addressee could see them. Gestures were smallest in the Computer condition ($M = 2.05$, $SD = .80$), followed by the Screen ($M = 2.24$, $SD = .51$), Web cam ($M = 2.37$, $SD = .65$), and FtF condition ($M = 2.60$, $SD = .38$). In the pairwise comparisons, there was a trend toward significance for the difference between the Computer and the FtF condition, $p = .07$.

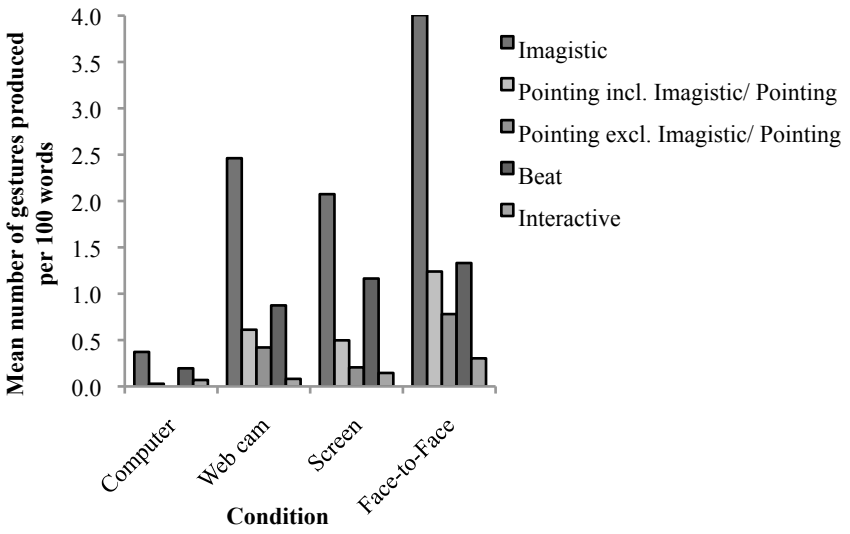

Figure 5: Mean number of gestures produced per 100 words per gesture type and condition.

The proportions of large and small gestures differed across conditions, as can be seen in Figure 6. Condition had a significant effect on the percentage of gestures involving shoulder movement, $F(3,32) = 4.04$, $p < .02$, $\eta_p^2 = .28$, see Figure 7. These gestures made up a significantly larger portion of the total number of gestures in the Web cam condition ($M = .16$, $SD = .14$) than in the Computer ($M = .03$, SD = .08), and Screen condition ($M = .01$, $SD = .03$). We found a trend toward significance, for the difference between the FtF ($M = .10$, $SD = .11$) and Screen condition, $p = .08$.

Gesture size and type For representational gestures, overall gesture size was very similar across conditions, ranging from $M = 2.69$, $SD = .25$ in the Screen condition, to $M = 2.98$, $SD = .51$ in the FtF condition, $F(3, 28) = .53$, $p = .67$. We found no significant main effect of condition on the size of imagistic, $F(3, 27) = 1.08$, $p = .37$, or pointing gestures, $F(2, 17) = 2.43$, $p = .12$. However, for pointing gestures, gesture size was significantly smaller in the Screen condition ($M = 1.67$, $SD = .82$) than in the FtF condition ($M = 2.76$, $SD = .68$). Combined imagistic/ pointing gestures were counted as imagistic in this analysis. We found no significant main effect of condition on the size of non-representational gestures, $F(3, 31) = 1.86$, $p = .16$. Yet pairwise comparisons showed that non-

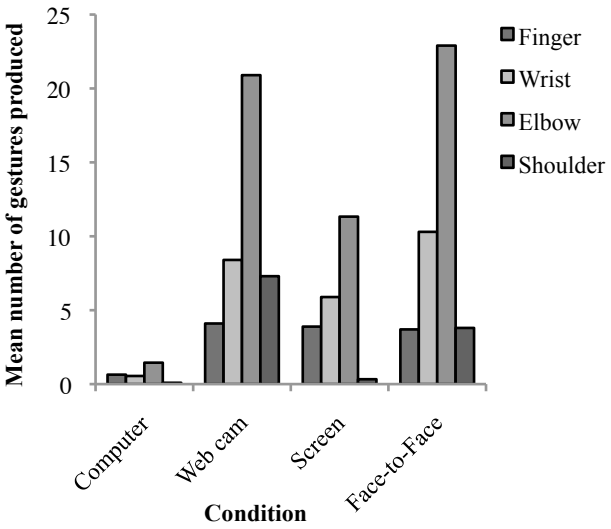

Figure 6: Mean number of gestures produced of each size in each condition.

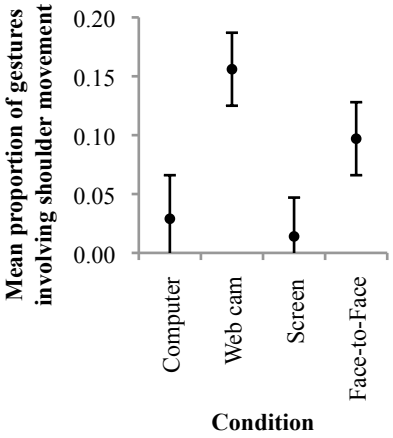

Figure 7: Mean proportion of gestures involving shoulder movement in each condition. Bars represent standard errors.

representational gestures were significantly smaller in the Computer (M = 1.34, SD = .47) than in the FtF condition (M = 1.87, SD = .36). No significant differences in the size of interactive gestures, $F(3, 14) = .40$, $p = .76$, and beats, $F(3, 31) = 1.42$, $p = .26$, were found when analyzing them separately.

Figure 8 gives an overview of the average size scores for the different gesture types for each condition. It must be noted though that some means are derived from very few data points, since some types of gesture were produced by only very few participants in some conditions. Figure 9 gives an overview of the mean number of gestures produced of each gesture type in each condition, and can help in interpreting Figure 8.

Number of words Condition had a significant effect on the total number of words used by participants, $F(3, 36) = 4.26$, $p < .02$, $\eta_p^2 = .26$, see Figure 10. In the Web cam condition (M = 841.80, SD = 373.93), significantly more words were used than in the Computer (M = 472.64, SD = 130.66) and FtF condition (M = 595.20, SD = 166.14).

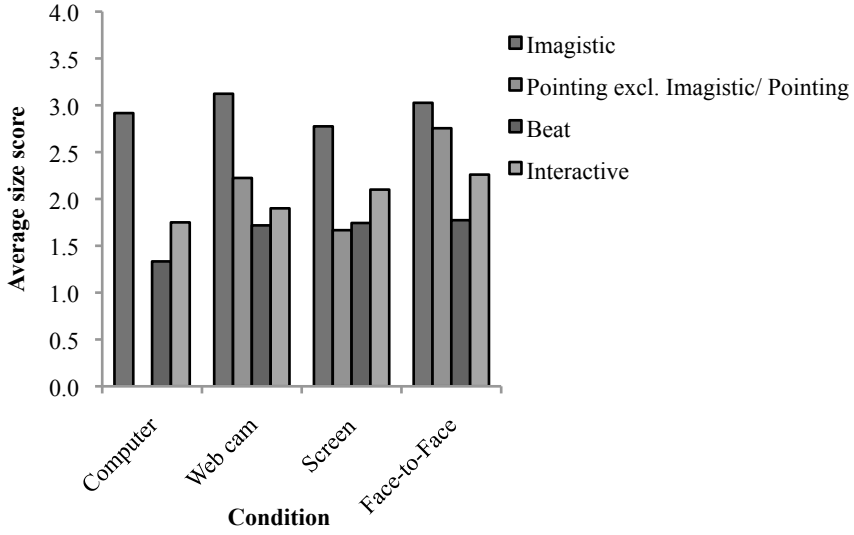

Figure 8: Average size score (1 = Finger, 2 = Wrist, 3 = Elbow, 4 = Shoulder) of gestures produced of each gesture type (Imagistic, Pointing, Beat, Interactive) in each condition.

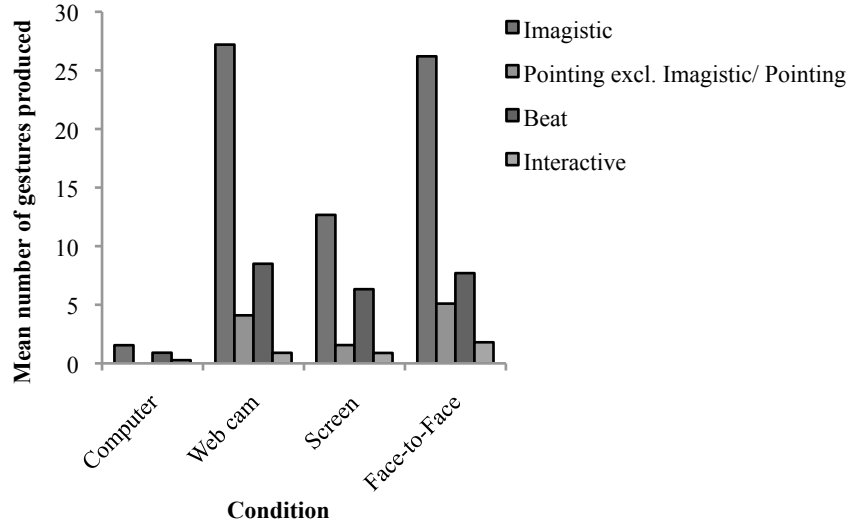

Figure 9: Mean number of gestures produced of each gesture type in each condition.

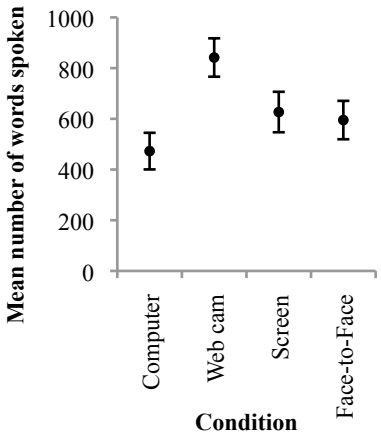

Figure 10: Mean total number of words spoken in each condition.
Bars represent standard errors.

Speech rate We found a significant effect of condition on the number of words spoken per second, $F(3, 36) = 4.92$, $p < .01$, $\eta_p^2 = .29$, see Figure 11. Speech was slower in the Computer condition ($M = 2.64$, $SD = .24$) than in the Screen ($M = 3.26$, $SD = .18$) and FtF condition ($M = 3.27$, $SD = .70$). Pairwise comparisons showed a trend toward significance for the difference between the Computer and Web cam condition ($M = 3.00$, $SD = .14$), $p = .07$.

Filled pauses No significant main effect of condition on the number of filled pauses (i.e. uhs) per word was found, $F(3, 36) = 1.82$, $p = .16$. However, pairwise comparisons showed that filled pauses were more frequent in the Web cam ($M = .10$, $SD = .04$) than in the FtF condition ($M = .06$, $SD = .04$).

Discussion

Participants who thought they were talking to an audiovisual summarizer produced fewer gestures than participants who thought they were talking to a human addressee, regardless of whether the addressee was in the same room or not and whether or not there was mutual visibility. Also, gestures produced by participants who believed that they were talking to the computer system were more frequently small (not involving shoulder movement) than the gestures produced by participants who thought they were talking to a human addressee

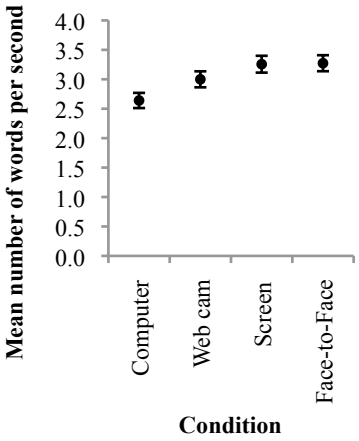

Figure 11: Mean number of words per second in each condition.
Bars represent standard errors.

through the web cam. So the (presumed) nature of the addressee, either human or artificial, clearly influenced gesturing.

The only difference between the Computer and Web cam condition was whether participants were told they were speaking to an audiovisual summarizer, or to another participant. Both were said to be in another room, so in both conditions the participant was narrating in front of a camera, without seeing or receiving any feedback from the addressee. Therefore, the difference in gesture rate and gesture size that we found between these two conditions can only result from speakers' mental representations of the addressee and this representation must include whether the addressee is artificial or not.

More words were used in the Web cam condition than in the Computer condition. Participants also spoke a little more slowly when they thought they were interacting with the computer system. Part of the difference in gesturing that we found between these two conditions could therefore result from differences in verbal behavior, rather than directly from differences in speakers' knowledge of the addressee's nature. However, the Web cam condition rather than the computer condition is the atypical one when looking at the number of words. The number of words used in the Computer condition did not differ significantly from the number of words used in the Screen and Face-to-Face condition, whereas the difference in gesture rate between the Computer

condition and these two conditions is striking. Descriptions in the Computer condition were generally detailed and elaborate, just like in the other conditions. We therefore think it unlikely that possible differences in the verbal behavior are the only source of the differences in the gestural behavior that we found. In addition, it would be hard for such a theory to explain why pointing gestures were almost completely absent in the Computer condition, while the same spatial content had to be expressed. Rather, we think that both the verbal and gestural modality were affected by the addressee being an artificial system or a human participant in another room.

Is the comparison between our Web cam and Computer condition a valid way of comparing human–human and human–machine communication? As can be seen from Figure 8 and 6a, the gestural behavior of participants in the Web cam condition was very similar to that of participants in the FtF condition. Similar patterns can be observed for the proportions of different gesture types and sizes. In this way, gesture production in the Web cam condition resembles that in the FtF condition. When looking at the average gesture rate (Figures 2 and 5), the Web cam condition is more similar to the Screen condition. Both these conditions show a lower gesture rate than the FtF condition. This suggests that both not seeing the addressee, or the absence of the possibility of dialogue, and not being seen by the addressee decrease gesture production. The comparison between the Computer and Screen condition shows that the very low gesture rate in the Computer condition does not just result from speakers not being seen by a human addressee. Neither can the low gesture rate be explained by the factor of not seeing the addressee or the lack of physical co-presence, since the Computer condition differed significantly from both the Web cam and the FtF condition. It thus seems that our design was indeed able to capture the difference between human–human and human–machine communication we were interested in.

The effect of mutual visibility on gesture production was replicated for the number of representational gestures per word, the number of pointing gestures per word, and the size of pointing gestures. For these variables we found significant differences between the Screen and FtF condition, as did earlier studies (e.g. Alibali, et al., 2001; Bangerter & Chevalley, 2007).

People behaved very differently toward the artificial system as compared to how they behaved toward people. This is contrary to the findings of Reeves and Nass (1996), who state that people behave toward ‘media’ as they would toward

a real person. It seems that speakers conveyed less information to the artificial system. It is unlikely that information was mostly transmitted through speech instead of through gestures when talking to the computer, since fewer and relatively fewer large gestures were produced, while participants did not use more words. Rather, it seems that less information was transmitted through both modalities. This corroborates well with the idea that people are less communicative when communicating to computers (Aharoni & Fridlund, 2007; Maes, et al., 2007). It would be interesting to do a comparative analysis of the verbal discourse to arrive at more clarity in this.

The differences we found in gesturing in different communicative settings can be explained by the idea that people make gestures for the benefit of their addressees. As explained under 'Present Study', we would find this explanation less believable if addressees are not sensitive to such differences. To test whether they are, we conducted two perception experiments, which will be described in the next section.

Perception study

It has been shown that addressees are able to process information from gestures (Beattie & Shovelton, 1999, 2002; Goldin-Meadow, 1999). However, in these studies information was directly related to the message a speaker was trying to convey, rather than to the communicative setting that a speaker was in. Chawla and Krauss (1994) found that observers could discriminate better than chance between spontaneous and rehearsed speech, both based on audio and audio-visual presentations. However, it remained unclear what cues observers had used in making their judgments.

With this study we want to determine whether observers are sensitive to differences in gesture production that result from differences in the communicative setting, especially the difference between addressing a human or an artificial addressee. At the same time, this perception study can be seen as a way to verify the gesture coding in our production experiment.

Experiment 1

In this experiment we asked observers to watch movie clips that were taken either from a setting with an artificial addressee (the Computer condition of our production study), or with a human addressee (the Screen condition of our production study). To separate the effect of gesturing from the effects of other visual cues, we measured the relative contributions of seeing the face and seeing the upper-body (including hands and arms) of the speaker.

Method

Design We used a between subjects design with three conditions. In condition 1, the 'Whole speaker condition', participants saw video clips in which the speaker's upper-body was fully visible. In condition 2, participants saw video clips in which the speaker's head was covered by a black rectangle (the 'Hands only condition'). And in condition 3, the 'Face only condition', participants saw video clips showing the head of the speaker only. In all conditions, the video clips were played without sound. After each video clip, participants were asked to judge whether the speaker was talking to a human or to an artificial addressee and to state on a binary scale whether they were certain or uncertain about their decision.

Participants Ninety first and second year students from Tilburg University and Eindhoven Technical University, all native speakers of Dutch, volunteered for this experiment. Most of them received half an hour of course credits for their participation.

Stimuli For this experiment we used 18 video clips from our production study: 9 of participants in the Screen condition, in which the story of an animated cartoon was retold to another participant (a confederate) who was seated behind an opaque screen, and 9 from participants in the Computer condition, in which participants retold the same story to a purported audiovisual summarizer. In both of these settings the speaker was seated in front of a camera.

From each video clip, 30 seconds were selected, starting from the point where the speaker began to describe the sixth episode of the cartoon, in which Sylvester builds a seesaw in order to catapult himself up to the window where

Tweety sits. This episode was chosen because it is very prone to elicit gestures. For the Whole Speaker condition, we used movie clips in which the speaker and all gestures were fully visible. Two different edited versions were then created, one such that everything was covered except the head of the speaker, for the Face Only condition, and one in which the head of the speaker was covered by a black rectangle, for the Hands Only condition.

Before the actual experiment started, there were two practice trials, for which video clips similar to the ones in the actual experiment were used. They were of a speaker in the Computer condition, and of a speaker in the Web cam condition of the production study.

Procedure Participants were randomly assigned to one of the three conditions. First, they read a written instruction and could ask the experimenter any questions they had. The instruction explained the task, but only stated that the participant had to indicate whether the speaker was talking to a human addressee or to an audiovisual speech recognition system. Details about the communicative setting, such as the difference in visibility (the computer could make use of video whereas the human addressee could not see the speaker) or co-presence (the computer was in another room, whereas the human addressee was in the same room) were not mentioned. Participants then did two practice trials, on which they did not receive any feedback. After the practice trials, the experimenter asked them again whether the task was clear and gave further instruction if necessary. Then followed the actual experiment.

Fragments were shown on a computer monitor. Half of the participants watched them in a certain random order, and the other half in reversed order. After each video fragment the screen turned black for seven seconds. On the black screen a sentence was shown in white, stating which fragment the participant was to fill out. This text disappeared after six seconds. The seven second pause was to be used by the participant to fill out on a paper sheet whether the speaker in the previous clip was talking to an audiovisual speech recognition system or to a human addressee, and whether the participant was certain or uncertain about this judgment (binary scale). After having judged all video clips, participants completed a brief questionnaire asking what features of the stimuli they had used in judging.

Results

Error rate The error rate refers to the proportion of movie clips that were judged incorrectly by a participant. We found a significant effect of condition on the average error rate, $F(2, 87) = 6.68$, $p < .01$, $\eta_p^2 = .13$. The error rate was significantly higher in the Face Only condition ($M = .34$, $SD = .12$) than in the condition in which participants could see the speaker entirely ($M = .22$, $SD = .10$), and in the condition where the face could not be seen ($M = .25$, $SD = .17$), see Figure 12. The latter two conditions did not differ significantly. The error rate was significantly below chance (.5) in all conditions. For the Whole Speaker condition: one-sample $t(29) = -15.57$, $p < .0001$, for the Hands Only condition: one-sample $t(29) = -8.15$, $p < .0001$, and for the Face Only condition: one-sample $t(29) = -7.11$, $p < .0001$.

We also found significant correlations between the number of gestures in our coding of the fragments and the number of participants who thought the speaker was talking to a human addressee (Whole: $r(28) = .88$, $p < .01$, Hands Only: $r(28) = .81$, $p < .01$).

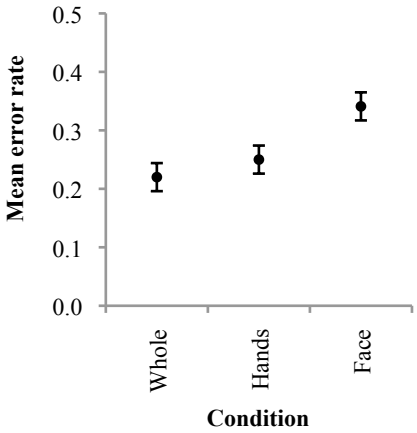

Figure 12: Mean error rate in each condition. Bars represent standard errors.

Discussion

The results of experiment 1 suggest that hand gestures are an important cue when judging whether a speaker is addressing a human addressee or a computer system. Participants could make this judgment reliably better than chance, even when they only saw the hands and upper-body of the speaker (without the face), and could not hear the speaker. They had the correct intuition that more hand gestures were produced toward a human, than toward the artificial addressee. The difference in gesturing that we found by analysis of the movie clips from our production experiment thus was confirmed by untrained observers, who could see parts of the movie clips only once.

In this first perception experiment, we compared movie clips from the Computer condition to movie clips from our Screen condition of our production experiment, rather than from our Web cam condition. Gesture rate, and overall gesture size did not differ significantly between these two conditions, although more very large gestures were produced in the Web cam condition. Also, in neither of these conditions did speakers receive visual feedback from the addressee. There was occasional auditory feedback in the Screen condition, but this was so rare that we trust it not to have had a major influence on our results, which is also indicated by the non-differing gesture rates. Nevertheless, one could argue that the differences that observers in the perception experiment made use of, resulted from a difference between the Computer and Screen condition of our production study other than the difference in the nature of the addressee. We therefore did a control experiment, in which movie clips of speakers from the Computer and Web cam condition were compared, to see whether participants could still reliably judge the nature of the addressee.

Experiment 2

Experiment 2 was similar to experiment 1, but this time we used movie clips from the Computer and Web cam condition of our production study. There was only one condition, in which participants could see the entire upper-body of the speaker. The instruction asked participants to judge whether a speaker was talking to an audiovisual speech recognition system in another room, or to a human addressee, who was watching them live on video from another room. Movie clips were played without sound, and were projected life-size onto a wall.

Sixty Master students from Tilburg University, all native speakers of Dutch, volunteered to participate in this experiment.

Results

The error rate (M = .33, SD = .08) was significantly below chance (.5), one-sample $t(59)= -16.87$, $p < .0001$. There was a significant correlation between the number of hand gestures in our annotation and the number of participants that thought a speaker was talking to a human addressee, $r(58) = 0.81$, $p < .001$.

Discussion

The results of our perception experiments clearly confirm that there are differences in gesture production when talking to a human addressee or to a computer system (even though the human addressees in the production experiment could not always see the speaker). More importantly, they show that observers are sensitive to these differences and have an intuition about how speakers gesture when talking to a human addressee or to an artificial system. When asked to explain the basis of their judgments afterward, most participants answered that they thought more gestures would be produced when talking to a human addressee, which is indeed the case.

Many participants also made comments on facial expressions. They expected speakers to be more vivid toward human addressees. Though gestures were the better cue for judging movie clips in experiment 1, we do not conclude that information from the face is less relevant to addressees. We did not inform viewers in experiment 1 that speakers could not see their addressee, or be seen by their addressee. Therefore, information from the face may have been misleading. Also, mutual visibility may influence facial expressions more than it does gesturing.

Even though in both studies participants performed better than chance, the error rate was lower in experiment 1 than it was in experiment 2. We think this may have to do with differences between speakers. Individual differences in gesture rate were relatively large among speakers from the Web cam condition. Apparently, some speakers matched the observers' expectations better than others. Participants expected speakers in the Web cam condition to gesture more than speakers from the Computer condition, but for some speakers this difference was quite small. This may have to do with the selection of the fragments from the speakers' narrations. We chose an episode in which

relatively many gestures were produced, which causes there to be relatively many gestures especially in the Computer condition, in which usually only a few gestures occurred throughout the entire narration. In addition, some speakers, especially in the Web cam condition, may have had more difficulty imagining their addressee than others.

General discussion

Our production study shows that just the speaker's idea of the nature of the addressee can be enough to influence gesture rate, the type of gestures produced, and the size of the produced gestures. Speakers gestured a lot more toward human addressees than toward a presumed audiovisual summarizer, they did not make pointing gestures toward the artificial addressee, and gestures that involved movement of the shoulder made up a larger portion of the gestures when talking to a human addressee through web cam than when talking to the artificial system. Since we found that people can largely refrain from gesturing, and do so spontaneously when asked to retell a story to a computer system, we conclude that gesture production is not a fully automated process and that it is related to the addressee.

Why would it be that people hardly produce gestures toward an audiovisual summarizer? One reason may be that information in gestures is largely redundant with information in speech. It could be that people do not expect a computer system to need such redundant information. Or perhaps gestures are not symbolic enough in nature, but rather relate to knowledge of the world too directly for speakers to expect the computer to benefit from them. Another reason may be that speakers did not feel the need to accommodate the artificial system as much as a human addressee. It has been found that speakers adapt less to an artificial addressee provided that it does not give feedback (Branigan, Pickering, Pearson, & McLean, 2010; Maes, et al., 2007). This may have caused speakers to be less informative in the gestural modality, but also they may have felt free to speak as slowly as they needed to the artificial system, thereby not needing gestures to "organize rich spatio-motoric information" (Kita, 2000, p. 163) or to facilitate word retrieval (Krauss, 1998). It would be interesting to measure the effect of time pressure on gesturing to test these hypotheses.

From the perspective that gestures are intended communicatively, the question remains open why the difference in gesturing is not more dramatic when people can or cannot be seen by their addressee. This may have had to do with the relative unresponsiveness of our addressees. Another possibility is that it was difficult for participants to apply their knowledge that the addressee could or could not see them. It has been shown for example that people do not always make optimal use of their knowledge of what the addressee can and cannot see when interpreting referential expressions (Keysar, Lin, & Barr, 2003). The small difference between the gesture rates in our Screen and Web cam condition somewhat points in this direction. One would expect speakers to gesture more frequently in the Web cam condition, in which they can be seen by their addressee, yet we found very comparable gesture rates for each gesture type in the Web cam and Screen condition.

If speakers indeed had problems applying their knowledge of the addressee, then the difficulty of the narration task may have further contributed to speakers not fully adjusting their behavior to the communicative setting. Most participants had some problems remembering parts of the animated cartoon they were retelling. Both processes: using one's knowledge of the addressee and remembering the story of the animated cartoon, may compete for the same cognitive resources. In a follow-up experiment, we manipulated the memory demands of the narration task, to observe whether participants adapt their (verbal and non-verbal) language production more to the communicative setting when doing an easier task, or whether they always gesture less when memory demands are lower (Mol, Krahmer, Maes, & Swerts, 2009).

A third possible explanation of our results is that gesturing is foremost a social activity. This social aspect may be a largely automated process that is simply not applicable when interacting with a computer system. This corroborates well with the idea of gestures having an interactive function (Bavelas, et al., 1992; Bavelas, et al., 2008). Yet it goes against the idea formulated by Reeves and Nass (1996), that people's responses to media are fundamentally social in nature. However, their studies did not avoid personalizing the computer by, for example, asking questions such as 'Did the computer help you well?', whereas we carefully formulated our instructions without attributing human actions, qualities or intentions to the audiovisual summarizer. The wording of such questions and instructions may influence the way participants think about an artificial system.

It has also been shown that the difference between gesturing on the phone and in a face-to-face situation is qualitative rather than quantitative (Bavelas, et al., 2008). Gesturing on the phone or to a person behind a screen may therefore serve a different purpose than does gesturing face-to-face. Still, our study shows that even this type of gesturing has something to do with interpersonal communication, besides the effect of dialogue, and may not be fully automated.

The effects of visibility on different gesture types that we found corroborate well with the results found in an earlier study (Alibali, et al., 2001). For representational gestures we found that significantly more gestures were produced in the Face-to-Face condition, in which speaker and addressee could see each other, than in the Screen condition in which they could not. This supports the hypothesis that representational gestures can be intended for the addressee. However, we did not find this difference between the Web cam and Screen condition. In the Web cam condition addressees were said to be able to see the speaker, but the speaker could not see the addressee and speaker and addressee were not physically co-present. One or both of these factors may influence the rate of representational gestures produced. For non-representational gestures, we found no significant difference between the Face-to-Face and Screen condition. However, we did find a difference between the Computer condition and the conditions with a human addressee, which may point to a communicative function of these gestures. In both the study by Alibali et al. and our study, large individual differences between speakers were found.

Like Bangerter and Chevalley (2007), we found an effect of visibility on the size of pointing gestures. Pointing gestures were larger in the Face-to-Face, than in the Screen condition. We also found that fewer pointing gestures were produced in the Screen condition than in the FtF condition and that no pointing gestures were produced toward the audiovisual summarizer. This supports the idea that pointing gestures are meant to be communicative and that their size is relevant to their meaning (Enfield, et al., 2007).

Our perception study has shown that gestures can be highly informative about the communicative setting that a speaker was in. Even when only seeing a speaker's gestures and not hearing the speaker, viewers could reliably judge whether that speaker had been talking to a human addressee or to an artificial system. This is consistent with a theory that speakers intend their gestures communicatively, as well as with a theory that speakers gesture mostly for themselves.

Conclusion

Whether the addressee is human or artificial can have an important influence on gesture production. People gesture less and produce a smaller proportion of gestures involving shoulder movement when narrating to an audiovisual summarizer, than when narrating to a human addressee. In addition, almost no pointing gestures were produced toward the artificial addressee. Just the speaker's mental representation of the nature of the addressee (either human or artificial) can be sufficient to influence the number and size of the gestures produced. We therefore conclude that gesture production is not a process that is fully automated in every communicative setting.

Given the size of the difference in gesture production that we found between narrating toward a human and an artificial addressee, it seems unlikely that gestures solely facilitate speech production. Rather, we think that some gestures are intended communicatively. However, part of the difference in gesturing that we found may relate to differences in verbal behavior.

A speaker's gestural behavior can convey information about the communicative setting that the speaker is in. It can reveal whether a speaker is talking to a human addressee or to a computer system. People are able to make this judgment better than chance from watching a speaker's hand gesture behavior alone.

Acknowledgements

We thank all participants. We gratefully thank Jan-Peter de Ruiter and Adam Kendon for the helpful and constructive comments on earlier versions of this work. We thank Carel van Wijk for his help in the statistical analysis, and Martin Reynaert and Lennard van de Laar for their technical support. Preliminary versions of this work were presented at ISGS 2007, AVSP 2007, and a workshop and master class on multimodal metaphors in Driebergen, 2007. Many thanks to the audiences and participants for their inspiring comments and questions on this research; in particular, we thank Sotaro Kita, Barbara Tversky, and Alan Cienky.

References

Aharoni, E., & Fridlund, A. J. (2007). Social reactions toward people vs. computers: How mere labels shape interactions. *Computers in Human Behavior, 23*, 2175-2189.

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language, 44*, 169-188.

Bangerter, A., & Chevalley, E. (2007). Pointing and describing in referential communication: When are pointing gestures used to communicate? In I. Van der Sluis, M. Theune, E. Reiter & E. Krahmer (Eds.), *Proceedings of the Workshop on Multimodal Output Generation* (pp. 17-28). Enschede: Universiteit Twente.

Bavelas, J., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse Processes, 15*, 469-489.

Bavelas, J., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language, 58*, 495-520.

Beattie, G., & Shovelton, H. (1999). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of Language and Social Psychology, 18*, 438-462.

Beattie, G., & Shovelton, H. (2002). An experimental investigation of some properties of individual iconic gestures that affect their communicative power. *British Journal of Psychology, 93*(2), 179-192.

Branigan, H. P., Pickering, M. J., Pearson, J., & McLean, J. F. (2010). Linguistic alignment between humans and computers. *Journal of Pragmatics, 42*, 2355-2368.

Chawla, P., & Krauss, R. M. (1994). Gesture and speech in spontaneous and rehearsed narratives. *Journal of Experimental Social Psychology, 30*, 580-601.

Cohen, A. A. (1977). The communicative functions of hand illustrators. *Journal of Communication, 27*(4), 54-63.

Cohen, A. A., & Harison, R. P. (1973). Intentionality in the use of hand illustrators in face-to-face communication situations. *Journal of Personality and Social Psychology, 28*, 276-279.

De Ruiter, J. P. (1998). Gesture and Speech Production. Unpublished Doctoral Dissertation. University of Nijmegen.

Dupont, S., & Luettin, J. (2000). Audio-visual speech modeling for continuous speech recognition. *IEEE Transactions on Multimedia, 2*(3), 141-151.

Enfield, N. J., Kita, S., & De Ruiter, J. P. (2007). Primary and secondary pragmatic functions of pointing gestures. *Journal of Pragmatics, 39*, 1722-1741.

Fisher, R. A. (1951). *The design of experiments*. Edinburgh: Oliver & Boyd.

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences, 3*(11), 419-429.

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining Math: Gesturing Lightens the Load. *Psychological Science, 12*(6), 516-522.

Goldin-Meadow, S., & Sandhofer, C. M. (1999). Gestures convey substantive information about a child's thoughts to ordinary listeners. *Developmental Science, 2*(1), 67-74.

Hadar, U. (1989). Two types of gesture and their role in speech production. *Journal of Language and Social Psychology, 8*(3-4), 221-228.

Hostetter, A. B., Alibali, M. W., & Kita, S. (2007). Does sitting on your hands make you bite your tongue? The effects of gesture inhibition on speech during motor descriptions. In D. S. McNamara & G. Trafton (Eds.), *Proceedings of the 29th annual meeting of the Cognitive Science Society* (pp. 1097-1102). Manwah, NJ: Erlbaum.

Hostetter, A. B., & Hopkins, W. D. (2002). The effect of thought structure on the production of lexical movements. *Brain and Language, 82*, 22-29.

Jacobs, N., & Garnham, A. (2007). The role of conversational hand gestures in a narrative task. *Journal of Memory and Language, 56*(2), 291-303.

Kendon, A. (1994). Do gestures communicate? A review. *Research on Language and Social Interaction, 27*(3), 175-200.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*(1), 25-41.

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and Gesture*. Cambridge: Cambridge University Press.

Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science, 7*, 54-60.

Maes, A., Marcelis, P., & Verheyen, F. (2007). Referential collaboration with computers: do we treat computer addressees like humans? In M. Swarz-Friesel, M. Consten & M. Knees (Eds.), *Anaphors in text: cognitive, formal and applied approaches to anaphoric reference* (pp. 49-68). Amsterdam: John Benjamins Publishing Company.

McCowan, I., Gatica-Perez, D., Bengio, S., Lathoud, G., Barnard, M., & Zhang, D. (2005). Automatic analysis of multimodal group actions in meetings. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(3), 305-317.

McNeill, D. (1992). *Hand and Mind: What gestures reveal about thought*. Chicago and London: The University of Chicago Press.

Melinger, A., & Levelt, W. J. M. (2004). Gesture and the communicative intention of the speaker. *Gesture, 4*(2), 119-141.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2009). Communicative gestures and memory load. In N. A. Taatgen & H. Van Rijn (Eds.), *Proceedings of the 31th Anual Conference of the Cognitive Science Society* (pp. 1569-1574).

Özyürek, A. (2002). Do speakers design their cospeech gestures for their addressees? The effects of addressee location on representational gestures. *Journal of Memory and Language, 46*(4), 688-704.

Reeves, B., & Nass, C. (1996). *The Media Equation, how people treat computers, televisions, and new media like real people and places*. New York: Cambridge University Press (CLSI Publications).

Chapter 3

Seeing and being seen: the effects on gesture production

Abstract
Speakers are argued to adapt their language production to their addressee's needs. For instance, speakers produce fewer and smaller hand gestures when interlocutors cannot see each other. Yet is this because speakers know their addressee cannot see them, or because they do not see their addressee? By means of computer-mediated communication we manipulated these factors independently. We found that speakers took into account what their addressee saw. They produced more and larger gestures when they knew the addressee could see them. Seeing the addressee increased gesture production only if the setting allowed for near-natural gazing behavior, which is not usually the case in mediated interaction. Adding this affordance resulted in gesturing being similar in mediated and unmediated communication.

This chapter is based on:
Mol, L., Krahmer, E., Maes, A., & Swerts, M. (In Press). Seeing and Being Seen: The effects on gesture production. *Journal of Computer-Mediated Communication*.

Introduction

Language use sometimes requires taking into account what another person can or cannot see. For example, when watching a documentary on Venice with a friend, you might ask your friend "have you ever been there?", where *there* refers to Venice. However, if your friend was in the same room, but working on her computer "have you ever been to Venice?" may be more appropriate. Because you know your friend is not watching the documentary, you may choose a more explicit reference. On the other hand, if you were asked by your friend, "have you ever been there?", while working on your computer, your knowledge of her watching a documentary on Venice may help in arriving at the correct interpretation. Yet would you do so correctly if you happened to be browsing a website on Berlin?

Language production and interpretation are argued to be adapted to our knowledge of another interlocutor, including what the other person knows about and sees (e.g. Grice, 1989). In this paper we focus on language production. Gesture and speech production can both be considered part of a speaker's language production (Kendon, 2004; McNeill, 1992). Therefore, one way of measuring to what extent speakers adapt their language production to their addressee is by looking at the hand gestures people spontaneously produce along with speech. It is well established that speakers produce fewer and different co-speech gestures when interlocutors cannot see each other (Cohen & Harison, 1973), such as on the phone (Bavelas, Gerwing, Sutton, & Prevost, 2008) or on both ends of an opaque screen (Alibali, Heath, & Myers, 2001). Thus, speakers seem to take into account what their addressee cannot see and thus cannot know about. Yet several empirical studies suggest that interlocutors tend to base their language production and interpretation on *their own* visual perspective, rather than that of their conversation partner (e.g. Keysar, Lin, & Barr, 2003; Wardow Lane, Groisman, & Ferreira, 2006). So do speakers truly take into account what their addressee sees?

Traditionally, in studies on hand gesture production, visibility has been manipulated symmetrically, such as by a screen that keeps the addressee from seeing the speaker, but at the same time keeps the speaker from seeing the addressee. We know that speakers tend to gesture less when visibility is obstructed in this way. Yet is this due to the addressee not seeing the speaker, or because of the speaker not seeing the addressee? Can we correctly apply our

knowledge of what the other sees to our language production, or do we tend to base it on our own observations?

To answer these questions, we need to somehow manipulate visibility asymmetrically, such that the speaker's perspective differs from the addressee's. Computer-mediated communication offers this possibility. With communication through web cams, we can separate the factors of seeing the addressee and being seen by the addressee, thus gaining important insights into what knowledge speakers apply to their language production.

To support the validity of this method, we need to draw the comparison between mediated and unmediated communication. Can we generalize from the results found in mediated communication to unmediated communication? What exactly is needed to make these two forms of communication optimally similar? Is it enough for interlocutors to have a live audiovisual presentation of each other? One possibly important difference between face-to-face interaction and communication through web cams is how readily interlocutors can interpret each other's eye gaze. We know that gaze and mutual gaze serve a variety of functions in unmediated interaction (Argyle & Cook, 1976; Kendon, 1967). In the second part of this paper we examine whether being able to interpret each other's eye gazing patterns influences speakers' language production in mediated communication.

Before describing how we used computer-mediated communication to study perspective taking, and presenting our findings on the importance of (mutual) gaze, we provide a brief overview of the literature on the communicative use of co-speech hand gestures, on when interlocutors have trouble taking into account each other's visual perspective, on the comparison of mediated and unmediated communication, and on the role of eye-gaze in interaction.

Background

Communicative co-speech gestures

When talking, many people move their hands and arms around, without the objective of directly manipulating their environment: they gesture. Several functions of such hand gestures have been identified, such as facilitating speech production (e.g. Hadar, 1989; Krauss, 1998), supporting learning (e.g. Goldin-Meadow, 2010), and aiding cognition (e.g. Chu & Kita, 2008; Melinger & Kita,

2007). In this paper, we focus on the communicative import of co-speech hand gestures. Numerous studies have addressed the communicative role of co-speech gestures. Some studies have shown that addressees gain information from gesture (e.g. Beattie & Shovelton, 1999; Cassell, McNeill, & McCullough, 1998; Chawla & Krauss, 1994; Cutica & Bucciarelli, 2008; Goldin-Meadow & Sandhofer, 1999; Mol, Krahmer, Maes, & Swerts, 2009). Additionally, other studies have rendered converging evidence that hand gestures are part of a speaker's communicative effort (Kendon, 2004). We will address this viewpoint below.

Kendon (1988) recognizes a continuum of how conventional and language-like hand gestures are. On the one end of this continuum are sign languages, in which signs have a conventional meaning and can be interpreted in the absence of speech. On the other end of the continuum is *gesticulation*. This is the production along with speech, of gestures that are not embedded in the grammatical structure of speech. For example, while saying "he went away", one could move one's arms back and forth along one's upper body, illustrating the manner of the action described. Alternatively, quickly wiggling one's down pointing fingers while moving the hand forward could illustrate the same event. Gestures at this end of Kendon's continuum are the most idiosyncratic gestures and their interpretation is highly dependent on the accompanying speech (Feyereisen, Van de Wiele, & Dubois, 1988).

These co-speech gestures are generally divided into several categories (e.g. McNeill, 1992). One broad distinction can be made based on whether a gesture depicts some of the content of the message a speaker is trying to convey, or whether it rather structures the conversation (e.g. Bavelas, Chovil, Lawrie, & Wade, 1992), or emphasizes certain parts of speech (e.g. Effron, 1941; Ekman & Friesen, 1969; Krahmer & Swerts, 2007). In this paper we focus on gestures that express some of the content a speaker is conveying, which are known as *illustrators* (Ekman & Friesen, 1969), or *representational gestures* (McNeill, 1992). Especially these gestures have been found to be produced differently by speakers in different communicative settings (e.g. Alibali, et al., 2001).

Effects of the communicative setting on representational gestures

When gestures have communicative potential, that is, when they can be seen by an addressee, they have been found to be larger (Bangerter & Chevalley, 2007; Bavelas, et al., 2008) and more frequently produced (Alibali, et al., 2001;

Bavelas, et al., 2008; Cohen, 1977). Yet does this imply that speakers take into account the addressee's visual perspective? In one study, Bavelas et al. (2008) asked participants to describe a picture of an elaborate dress. When participants interacted face-to-face, their gestures about the dress were full sized, as of an actual dress one could wear, whereas when interaction took place over the telephone, gestures were only the size of the dress in the picture. Although this study clearly illustrates speakers' sensitivity to the communicative context, we can not be sure that speakers were adapting to what their addressee saw, since gesturing based on what they themselves saw would result in the same behavior.

Alibali et al. (2001) found that when speakers were asked to retell an animated cartoon to an addressee, they produced representational gestures more frequently in a face-to-face setting than when speaker and addressee were separated by an opaque screen. Again, this shows speakers' sensitivity to the environment. Yet it does not tell us whether speakers were taking into account their addressee's visual perspective, since whether or not the addressee saw the speaker corresponded with whether or not the speaker saw the addressee. Therefore, speakers may have based their gesturing on their own visual perspective. Using a similar narration task, Jacobs and Garnham (2007) found that gestures were produced more frequently when a speaker *knew* that information was new to the addressee as well as when the addressee appeared attentive and interested. This suggests that speakers do take into account listener needs, but again these needs were also readily visible to the speaker. (Speakers either retold the same cartoon twice to the same addressee or to two different addressees, and either saw an interested or a less interested addressee.) This also holds for a study by Özyürek (2002), which showed that speakers produce their gestures differently, depending on where the addressee is located relative to them. Although speakers were shown to change the orientation of their gestures based on the addressee's location, we do not know if this resulted from the change in their own visual perspective or in that of their addressee.

As described in the previous chapter, in an earlier study (Mol, et al., 2009), we manipulated whether speakers thought to be talking to a human addressee or to an audiovisual speech recognition system. In this study, contrary to the aforementioned studies, the environment of the speaker was exactly the same across conditions. The only difference was in the preceding instruction. The speaker was seated alone in a room in front of a camera and was either told that the audiovisual output of this camera was sent to another participant, or to an

artificial system. Speakers were found to gesture more frequently and produce more large gestures when they thought to be addressing a human addressee. This time, the difference in gesturing could only be caused by a different belief about the addressee. Yet it remains unknown whether speakers apply more of their knowledge about the addressee to their language use than just whether the addressee is human.

Although the above-mentioned results all point in the direction that speakers do apply their knowledge of their addressee to their hand gesture production, and thereby to their language production, these results leave open the possibility that the actual application of such knowledge is very limited and that speakers mostly use an egocentric perspective. That is, they may base their language production on what they themselves see. We therefore turn to some studies that have shown people's difficulty in applying knowledge about their addressee's visual perspective to their language use.

Taking into account what an interlocutor sees

Keysar, Lin, and Barr (2003) have shown that people tend to make 'mistakes' in their interpretation of speech, when arriving at the correct interpretation requires taking into account what a speaker can and cannot see. By studying participants' eye movements, they found that addressees mistakenly considered objects that they knew a speaker could not see as possible referents of speakers' referring expressions. Wardlow Lane, Groisman, and Ferreira (2006) found similar results for reference production. In their study, a speaker had private visual access to an object that only differed from the target object in size. Even though the addressee could not see this competing object, speakers often included a contrasting adjective, such as 'small', in their reference to the target object, despite this not being informative to the addressee. Surprisingly, they did so even more when instructed to conceal their private information from the addressee. Note that the contrasting adjective provides a cue to the properties of the object that was hidden from the addressee. Therefore, it seems that speakers have difficulty in applying their knowledge of what their addressee sees to their speech production. Similar difficulties may affect speakers' hand gesture production.

Computer-mediated vs. unmediated interaction

Mediated communication can help us resolve the issue of whether speakers employ an egocentric perspective or not, when it comes to adapting their co-

speech gestures to a communicative setting. Yet can mediated communication be representative of face-to-face interaction? Social presence theory (Short, Williams, & Christie, 1976) proposes that the extent to which social presence is experienced in mediated communication depends on the affordances offered. The more affordances available, the more warmth and affection interlocutors will experience and express. Social information processing theory (Walther, 1992) adds that interlocutors can also adjust both their motives and their communicative efforts to a medium, such that mediated communication does not necessarily fall short of face-to-face interaction when it comes to experienced presence. For example, Walther, Slovacek, and Tidwell (2001) have shown that seeing a picture of the addressee promotes affection and social attraction in short-term mediated interaction, but that this is not true for long-term mediated interaction. Given ample time, the highest levels of affinity were established through a text-based medium.

Consistent with this approach, Brennan and Lockridge (2006) use the grounding framework to describe how communication is affected by mediation: "The grounding framework conceptualizes mediated communication as a coordinated activity constrained by costs and affordances (Clark & Brennan, 1991)", p. 1. From this perspective, the more the costs and affordances of mediated communication resemble the costs and affordances of face-to-face interaction, the more similar the two will be. For example, Brennan and Ohaeri (1999) argue that mediated written conversation can be less polite compared to spoken interaction, because the production costs of politeness are higher when typing than when speaking. This in turn could lead to interlocutors perceiving each other differently, rather than these different perceptions resulting from mediation directly. From these frameworks, we can infer that both being able to see the addressee and being seen by the addressee will result in mediated communication being more similar to face-to-face interaction.

Communication through desktop video-conferencing, such as with Skype, offers many of the affordances available in face-to-face communication. The use of web cams and microphones allows speakers to see and hear each other almost real time, even though they are in different locations. Would this result in interlocutors behaving the same way as in face-to-face interaction? Isaacs and Tang (1994) observed interactions between technical experts that took place over the phone, through desktop video-conferencing, or face-to-face. They found that the experts indeed used the visual modality in video-conferencing much like they

did in face-to-face communication. "Specifically, participants used the visual channel to: express understanding or agreement, forecast responses, enhance verbal descriptions, give purely nonverbal information, express attitudes through posture and facial expression, and manage extended pauses", p. 65. However, Isaacs and Tang also listed some differences between video-conferencing and face-to-face communication, for example, managing turn-taking, having side conversations, and pointing to objects in each other's space were more difficult in video-conferencing.

One difference in affordances between video-conferencing and face-to-face interaction is the availability and interpretability of information from gaze. For example, when interlocutors are not co-present and the physical environment is not shared, the direction of each other's gaze cannot readily be interpreted. When using a web cam, it can even be misleading. Since the image of the conversation partner and the web cam are not in the same location, looking at the web cam means not looking at the other interlocutor. Yet when someone looks into the camera, their image misleadingly appears as though they are looking at the person watching the image. To what extent do the availability of an interlocutor's gaze and the ease with which it can be interpreted influence language production? Can the difference in the availability of information from gaze account for some of the differences found between communicating face-to-face and by means of video-conferencing?

Using information from others' gaze

We know that gaze and mutual gaze serve many functions in unmediated interaction (Argyle & Cook, 1976; Kendon, 1967). Among other functions, gaze can be used to infer if the other person is attending and whether a message is understood, as well as to solicit such signals from the conversation partner. It has also been found that when speakers gaze at their own gestures, addressees are more likely to fixate on these gestures as well, and are also more likely to retain information from these gestures (Gullberg & Kita, 2009). Thus, speakers may use gaze as a way to direct their addressee's attention to their gestures.

Hanna and Brennan (2007) found that addressees use speakers' eye gaze to disambiguate referring expressions. Addressees could do so both when a speaker's visual perspective matched their own perspective and when it was a mirror image, showing that they could map the speaker's visual perspective onto their own. Brennan, Xin, Dickinson, Neider and Zelinsky (2008) found that

participants were able to benefit from seeing another participant's gaze indicated on their screen, when performing a simple collaborative search task. Seeing each other's gaze represented by a cursor on the display was shown to allow for a more optimal division of labor than did talking to each other. This shows that participants were able to adapt their behavior, based on their knowledge of where their partner was looking.

These studies show that people can sometimes benefit from observing other people's gaze when communicating or co-operating, both in unmediated and mediated settings. They also show that the interpretation and production of gaze can be adapted to a task or a medium. How important is this factor when it comes to the difference between mediated and unmediated communication? Do interlocutors behave differently dependent on whether gaze is easily interpreted or not, or do they interpret gaze correctly independent of the effort involved, resulting in similar communicative behavior (including gesture production)?

Present study

Our present study consists of three experiments. First, we investigate whether the fact that speakers produce fewer hand gestures when interlocutors cannot see each other is due to the speaker not seeing the addressee, to the addressee not seeing the speaker, or to both of these factors. We do so by asking participants to perform a narration task in one of four settings, in which we independently manipulate visibility of the speaker and addressee, by means of communication through web cams. Second, we investigate how important it is for speakers to be able to readily interpret their addressee's gaze. For this we make use of a newer video-conferencing technique: the Eye-Catcher (GreenEyes, 2007). This device enables interlocutors to interpret each other's gazing behavior more readily than when using web cams. We measure how this affects gesture production. Third, we test what the differences in speakers' production behavior due to this difference in the interpretability of gaze, mean to naive observers.

Study 1: Seeing and being seen through web cams

In order to manipulate visibility asymmetrically, we make use of video-conferencing through web cams. Our communication task is chosen such that the differences found by Isaacs and Tang (1994) between video-conferencing and

face-to-face interaction are minimized. There are only two interlocutors, so there is no possibility of having side conversations. We use a task in which a speaker retells an animated cartoon to an addressee, who is instructed not to interrupt (after Alibali, et al., 2001). Therefore, there is little need for coordinating turns. Also, this task does not relate to the physical environment of either the speaker or the addressee, so there is no need to point at real objects in the environment. Therefore, for this task, the costs and affordances of video-conferencing are a close match to face-to-face interaction. Hence, we expect that manipulating mutual visibility will have similar effects in our mediated settings as it does in unmediated settings.

The use of gesture production as a dependent variable enables us to measure how participants' communicative behavior is influenced by the communicative setting, rather than how they subjectively experience it. We can look at both the frequency and the quality of the gestures produced. Both these factors have been related to communicative effort, and speakers are known to adjust these aspects of their communicative behavior to whether or not there is mutual visibility. Therefore, gesture is a suitable measure for determining if speakers tend to base their language use on their own visual perspective or if they correctly apply their knowledge of the addressee's visual perspective.

Seeing the addressee could influence gesture production for several reasons. One reason is that speakers may base their gesturing on their own visual perspective, and will therefore gesture as though there is mutual visibility when in fact they can only see the addressee. This would mean that speakers gesture more when they can see the addressee, regardless of whether the addressee can see them. If speakers solely use their own visual perspective, this would also mean that when speakers cannot see their addressee, they will produce an equal number of gestures, irrespective of their knowledge of whether the addressee can see them.

Another reason why seeing the addressee may influence gesture production is because of the signals the speaker receives from the addressee. In this case, seeing the addressee may elicit more gestures if it motivates speakers to put in more communicative effort, for example because they experience a higher degree of social presence (Short, et al., 1976) or affinity (Walther, et al., 2001), or because the addressee seems more interested (Jacobs & Garnham, 2007). Yet receiving cues from the addressee may also reduce gesture production, especially when the addressee's gaze is hard to interpret, which may cause the addressee to

appear inattentive. Regardless of its direction, this effect would be independent of the effect of being seen by the addressee.

Being seen by the addressee can only influence gesture production directly if speakers correctly apply their knowledge of the addressee's visual perspective, thereby possibly overriding or replacing an egocentric perspective. If speakers indeed base their language production on their knowledge of the addressee's perspective rather than their own visual perspective, this would mean they gesture more when they can be seen by the addressee, rather than when they see their addressee.

Method

Design We used a 2 x 2 between participants design in which we manipulated whether or not the speaker could be seen by the addressee and whether or not the speaker could see the addressee, see Figure 1. In all conditions speaker and addressee could hear each other.

Participants Thirty-eight (21 female) native Dutch speakers, all students from Tilburg University, participated in this study as part of their first year curriculum. Two participants were excluded from our analysis (see Analysis). The remaining 36 participants (20 female) had a mean age of 22, range 18 - 33.

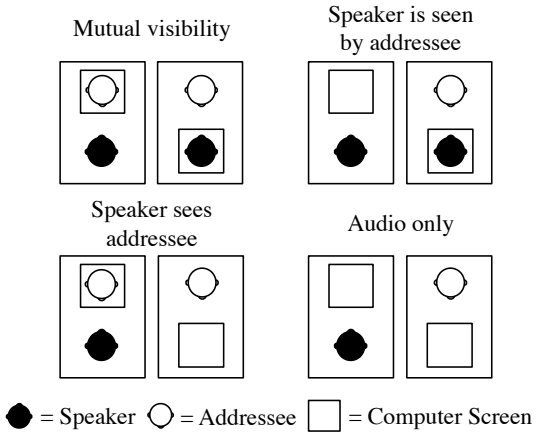

Figure 1: Schematic overview of the settings. Speaker and addressee are seated in separate rooms and can either see the other interlocutor on a monitor or not.

The addressee was a female confederate, who also was a student from Tilburg University.

Procedure The experimenter received the participant and the confederate in the lab and assigned the role of speaker to the participant and the role of addressee to the confederate. Narrators were asked to retell the story of an animated cartoon (*Canary Row* by Warner Bros) to the addressee. After reading the instructions, participants could ask any remaining questions. The confederate always posed a clarification question. The narrator's instructions stated that the addressee had to summarize the narration afterward and explained that the narrator was videotaped to enable comparison of the summary and the narration.

When all was clear the narrator was seated behind a table with a computer screen on it, which in some settings showed a live video-image of the addressee (full screen), and in the remaining settings showed just the interface of the video-conferencing application we used (Skype). If the addressee was shown, the entire upper body of the addressee was visible, rather than just the face. The computer screen was connected to a computer, which also had a web cam connected to it. Behind the table stood a tripod, which held the web cam and a digital video camera. The position of the web cam was such that the entire upper body of the speaker was captured. On the wall behind the video camera were eight stills from the animated cartoon, one from each episode, as a memory aid for the narrator and to elicit more structured and hence more comparable narrations.

The experimenter took the addressee to another room with a similar setup (but without the stills) and established a connection between the two computers over the internet, using Skype. Sound and video were both captured by the web cams and sound was played back through speakers. To familiarize the participants with the setting, sound was tested by the narrator and addressee talking to each other and if applicable, the video-image was tested as well. The connection was then suspended temporarily, while the narrator was left alone to watch the animated cartoon on a different computer. When the cartoon had finished the experimenter re-established the connection, and seated the narrator behind the camera. In conditions where the addressee could see the narrator, narrators were briefly shown what the addressee saw. In the remaining conditions, the experimenter repeated that the addressee could not see the narrator. The experimenter then started the video recording and left the room.

When the narrator was done telling the story, participants completed a questionnaire, which included questions on how the communicative setting had been experienced, how interested the addressee had appeared, and whether any deception was suspected. Meanwhile, the addressee ostensibly wrote a summary on yet another computer in the lab room. None of the participants had suspected any deception. After filling out the questionnaire, they were fully debriefed. All of the participants gave their informed consent for the use of their data, and if applicable for publishing their photographs.

During the narration, the confederate refrained from interrupting, laughing, etc. When necessary, minimal feedback was provided verbally. She always gazed somewhere near the web cam capturing her, independent of whether she could see the speaker.

Transcribing and coding We transcribed each speaker's narration from the videotape. Filled pauses, such as 'uh' were included in the transcription. A Perl script was used to compute the number of unique words in each narration.

We coded all hand gestures produced by speakers. Based on the gesture's shape and the accompanying speech, we coded whether a gesture depicted some of the content of the animated cartoon, or whether it was about the current conversation, e.g. placing emphasis. In our analysis we focus on the former category, which we refer to as *representational gestures*. Figure 2 depicts two examples of our coding. In the scene on the left, the speaker imitates a hitting motion while talking about someone hitting. In the scene on the right, the speaker refers to the main character and briefly moves his fingers up and down.

We also coded the size of each gesture. Gestures that were produced using only the fingers received a score of 1. If the wrist was moved significantly the gesture received a score of 2. Gestures that also involved significant movement of the elbow or lower arm received a score of 3, and gestures in which the upper arm was also used in a meaningful way, or that involved movement of the shoulder received a score of 4. This way, an average gesture size was computed for each participant.

Statistical analysis Analyses were done using a 2 x 2 ANOVA, with factors *speaker is seen by addressee* (levels: yes, no) and *speaker sees addressee* (levels: yes, no). The significance threshold was .05 and we report partial eta squared as a measure of effect size. As dependent variables for participants' gestures, we

Figure 2: Left: example of a representational gesture (depicting hitting), Right: example of a non-representational gesture (placing emphasis while referring to a character).

use the number of representational gestures produced per minute (*gesture rate)* and the average size of representational gestures. We use the mean gesture rate rather than the mean total number of gestures produced, to control for any differences in the duration of the narrations between participants. Twoparticipants were excluded from the analysis, because they deviated more than 2 standard deviations from the mean gesture rate in their condition. As a result, there were 9 participants in each condition. Inclusion of these two participants did not affect the significant effects found, but did reduce the significance of the overall model.

Although our focus is on participants' hand gestures, we also report some general measures of participants' speech. This addresses the question of whether different behavior in gesturing follows from the verbal narrations being much different across settings, rather than it being a direct result of the communicative setting. As global measures of the content of the narrations, we report the total number of words produced and the ratio of the number of unique words divided by the total number of words (*type-token ratio*). In addition, we report the number of words per second (*speech rate*) and the number of filled pauses per 100 words, as measures of fluency. To exclude a possible confounded effect of speech rate, we used the speech rate as a covariate in all our ANOVAs on gesture data in this and the following studies. Throughout the entire paper we report all means before this correction.

Results

Gesture rate Analyses of the number of representational gestures per minute, shown in Figure 3, revealed a main effect of the speaker being seen by the addressee, such that speakers gestured more frequently when they were seen (M = 5.66, SD = 5.82) than when they were not (M = 2.58, SD = 2.03), $F(1, 32) = 4.25$, $p < .05$, $\eta^2_p = .13$. The main effect of the speaker seeing the addressee showed a trend toward significance, such that speakers gestured less frequently when they saw their addressee (M = 2.75, SD = 3.35) than when they did not (M = 5.49, SD = 5.27), $F(1, 32) = 3.74$, $p = .06$, $\eta^2_p = .11$. The two factors did not interact, $F(1, 32) < 1$.

Gesture size Analyses of the average size of representational gestures, shown in Figure 4, revealed a trend toward significance for the main effect of the speaker being seen by the addressee, such that speakers' gestures were larger when speakers were seen (M = 3.03, SD = .73) than when they were not (M = 2.53, SD = .78), $F(1, 29) = 2.93$, $p = .10$, $\eta^2_p = .09$. When including non-representational gestures, being seen exerted a main effect on gesture size, $F(1,31) = 4.59$, $p < .05$, $\eta^2_p = .13$. The speaker seeing the addressee did not exert a main effect on the gesture size, $F < 1$, and the two factors did not interact, $F < 1$.

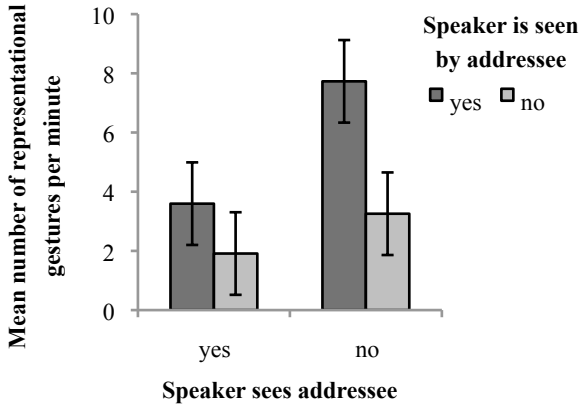

Figure 3: Mean gesture rate depending on whether the speaker could be seen by the addressee (separate columns) and whether the speaker could see the addressee (x-axis). Bars represent standard errors.

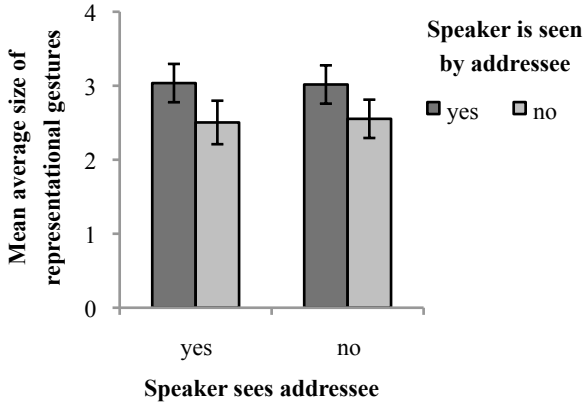

Figure 4: Mean average size of representational gestures, depending on whether the speaker could be seen by the addressee (separate columns) and whether the speaker could see the addressee (x-axis). Bars represent standard errors.

Speech Neither the speaker being seen by the addressee, nor the speaker seeing the addressee exerted a main effect on the total number of words used, the type-token ratio, or the number of filled pauses per 100 words. The speaker being seen by the addressee exerted a main effect on the number of words per second, such that speakers spoke faster when they were seen (M = 2.99, SD = .24) than when they were not (M = 2.73, SD = .45) , $F(1, 32) = 4.61$, $p < .05$, $\eta^2_p = .13$. The speaker seeing the addressee did not exert a main effect on the speech rate, $F < 1$. The two factors did not interact, $F(1, 32) = 1.05$, $p = .31$.

Perceived interest Analyses of the extent to which speakers perceived the addressee as disinterested, shown in Figure 5, revealed an interaction between the factors being seen by the addressee and seeing the addressee, $F(1, 31) = 11.87$, $p < .01$, $\eta^2_p = .28$: Speakers agreed to the statement that the addressee was disinterested more when they could see the addressee, but the addressee could not see them. The main effect of the speaker seeing the addressee showed a trend toward significance, such that speakers agreed to the statement more when speakers could see the addressee (M = 3.53, SD = 1.42) than when they could not (M = 2.89, SD = 1.28), $F(1, 31) = 3.09$, $p = .09$, $\eta^2_p = .09$.

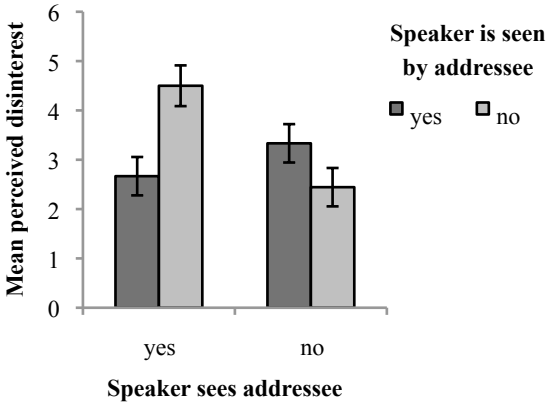

Figure 5: Means of speakers' answer to the statement "The addressee was disinterested" on a 7 point scale, 1 = completely disagree, 7 = strongly agree.
Bars represent standard errors.

Other Neither the speaker being seen by the addressee, nor the speaker seeing the addressee exerted a main effect on the duration of the narration. There was no significant correlation between the speech rate and the gesture rate, $p = .13$.

Discussion

Representational gestures were produced more frequently when speakers knew their addressee could see them. This was true both when speakers saw the addressee and when not, but especially the difference when speakers could not see their addressee is striking. This clearly shows that it was speakers' belief of being seen by the addressee that increased gesture production. In addition, gestures tended to be larger when speakers knew their addressee could see them, independent of whether speakers saw their addressee or not. These results clearly show that speakers applied their knowledge of the addressee's visual perspective to their gesture production, rather than solely using their own visual perspective.

Other than in previous work (Alibali, Kita, & Young, 2000; Bavelas, et al., 2008; Cohen, 1977; Jacobs & Garnham, 2007; Özyürek, 2002), our results cannot be explained by observable changes in the environment of the speaker. Our study therefore supports the interpretations of these earlier studies in terms of audience design. Speakers are able to adapt their gesturing to their (knowledge of their) addressee.

Participants' verbal narrations in each setting were similar, as can be seen from their length in words, the variation in vocabulary, their duration in time, and the frequency of filled pauses. Interestingly, participants spoke faster when they could be seen. It therefore seems that speakers also applied their knowledge of the addressee's perspective to their speech production. Perhaps speakers have an intuition that their speech can be understood more easily when they can be seen. It has been shown that visual information can indeed aid speech interpretation (e.g. Sumby & Pollack, 1954).

When speakers could see their addressee, they tended to produce fewer gestures than when they could not. At first sight it may seem surprising that speakers did not gesture more frequently when they saw their addressee, but the video-image of the addressee may have been confusing. The confederate addressee always gazed somewhere near the web cam capturing her, regardless of whether she could see the speaker or not. This somewhat unnatural gazing behavior may have been interpreted as the addressee being less interested. The results of our offline measure, a questionnaire, support this. After the narration task, the addressee was rated as less interested when the speaker could see the addressee but the addressee could not see the speaker. Speakers are known to gesture less when the addressee appears uninterested or inattentive (Jacobs & Garnham, 2007). Interestingly, this did not affect gesture size. Once a gesture was produced, it was produced larger when it was visible to the addressee, independent of how interested the addressee appeared to be.

Because of the effects of *being seen by the addressee* and *seeing the addressee* acting in opposite directions, the gesture rate in the setting in which speaker and addressee could see each other is surprisingly low compared to the setting in which only the speaker could be seen, and perhaps not quite representative of face-to-face interaction. Our second experiment addresses this issue.

Study 2: Eye-Catcher

One important difference remains between our setting with mutual visibility through web cams and a face-to-face setting: the usability of gaze. In study 2, we investigate the effect of the possibility for mutual gaze and the cost of interpreting gaze on language production. We do so by using a new technology

for mediated communication: the Eye-Catcher, which enables interlocutors to look at each other's video-image *and* straight into the camera at the same time. This way, interlocutors can both look at each other simultaneously and appear to be looking at each other as well. This is not possible in communication through web cams, where giving the impression of looking at the other interlocutor requires looking into the camera, while seeing the other interlocutor requires looking at the screen.

Note that for the task we use, there is not much difference between using a web cam and using Eye-Catchers when using one-way visibility. In this case, either the addressee has nothing of interest to look at and thus it is of little use to the speaker to be able to interpret the addressee's gaze, or the speaker does not see the addressee and thus cannot make use of the addressee's gaze. Therefore, we use the Eye-Catcher in a setting with mutual visibility only. In such a setting, the addressee's gaze can inform the speaker of what the addressee is looking at. By comparing the Eye-Catcher setting to the Web Cam setting with mutual visibility, we can see how the availability of information from gaze affects language production.

If the availability of mutual gaze affects language production, we can re-examine the effects of being seen by the addressee and seeing the addressee, replacing the data from the Web Cam setting with mutual visibility with the data of the Eye-Catcher setting. This may provide a closer match with unmediated communication. Also, we can compare the effect of mutual visibility in mediated and unmediated communication, to see if the results obtained with mediated communication are likely to generalize to unmediated communication. To draw this comparison, we make use of the data of an earlier study (Mol, et al., 2009), in which we used the same paradigm of retelling a cartoon in two unmediated settings. In the Face-to-Face setting, speaker and addressee were seated in the same room facing each other, such that visibility was unimpaired. In the Screen setting, interlocutors sat in the same room but on either side of an opaque screen, such that they could not see each other.

Method

The method of our second study was the same as in our first study, except that this time we used Eye-Catcher technology instead of communication through web cams. The Eye Catcher consists of a one-way mirror, in which a screen is reflected. Behind the one-way mirror is a camera. This way, the person in front

Figure 6: The Eye Catcher seen from the front, with a video image of the addressee.

of the Eye-Catcher can be captured while watching the reflected screen. We used two connected Eye-Catchers, such that the image captured by one Eye-Catcher's camera was shown on the screen of the other Eye-Catcher. This way, it appears as though interlocutors can look each other in the eyes, see Figure 6.

Through the Eye-Catchers, the narrator and the (confederate) addressee were able to see and hear each other and there was a possibility for mutual gaze. They were each seated in front of a table with an Eye-Catcher on it, at the same angle and distance, such that the setting was maximally symmetrical, enhancing the interpretability of gaze. The confederate's gazing was natural and her other back-channeling behavior was restricted in the same way as before.

Participants Nine (6 female) native Dutch speakers, all students from Tilburg University, participated in this study as part of their first year curriculum. They had a mean age of 21, range 19 - 26. In the earlier study we used for comparative analyses, 19 native Dutch students from Tilburg University participated. In the Face-to-Face setting, there were 10 (8 female) participants with a mean age of 19, range 17 - 21. In the Screen setting, there were 9 (7 female) participants with a mean age of 18, range 18 – 19.

Transcribing, coding & statistical analysis We transcribed participants' speech and coded their hand gestures in the same way as before. First, we compare the Eye-Catcher setting to the Web Cam setting with mutual visibility in an independent samples *t*-test. This way, we can see if and how the Eye-Catcher's

affordance of mutual gaze affects gesture and speech production. Second, we repeat our ANOVA with factors *speaker is seen by addressee* (levels: yes, no) and *speaker sees addressee* (levels: yes, no), with the data of the Eye-Catcher setting replacing the data of our previous setting with mutual visibility through web cams. The one-way visibility settings were not replaced in this analysis. As explained earlier, there is no relevant difference between using Eye-Catchers and using a web cam in one-way visibility settings. We also perform an ANOVA with factors mutual visibility (levels: yes, no) and mediation (levels: yes, no), comparing the mediated Audio Only setting of experiment 1 and the Eye-Catcher setting, to the unmediated Face-to-Face setting and the Screen setting of our earlier study (Mol, et al., 2009).

Results

Comparing the Web Cam and Eye-Catcher settings

Gesture Representational gestures were produced more frequently in the Eye-Catcher setting (M = 14.25, SD = 8.08) than in the Web Cam setting (M = 3.60, SD = 4.26), $t(16) = 3.50$, $p < .01$. The gesture size was comparable in both conditions, $t(16) = .26$, $p = .80$.

Speech The total number of words produced, the duration of the narration, the type-token ratio, and the number of filled pauses per 100 words did not differ significantly across the two settings. The speech rate was also comparable in the Web Cam (M = 2.92, SD = .29) and Eye-Catcher setting (M = 3.01, SD = .45), $t(16) = -.46$, $p = .74$.

Analysis with factors: speaker is seen by addressee, speaker sees addressee
Since the gesture rate was much different in the Web Cam and Eye-Catcher setting, we repeat our previous analysis of experiment 1, with the data of the Eye-Catcher setting replacing the data of the Web Cam setting with mutual visibility.

Gesture rate Analyses of the number of representational gestures per minute, shown in Figure 7, revealed a main effect of the speaker being seen by the addressee, such that speakers gestured more frequently when they were seen (*M*

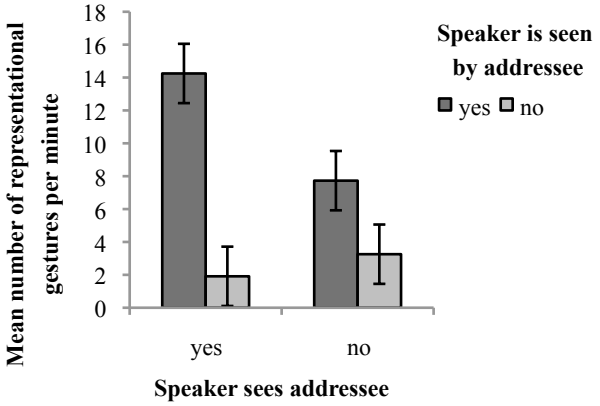

Figure 7: Mean gesture rate depending on whether the speaker could be seen by the addressee (separate columns) and whether the speaker could see the addressee (x-axis). Bars represent standard errors.

= 11.0, *SD* = 7.92) than when not (*M* = 2.58, *SD* = 2.03), $F(1, 31) = 15.34$, $p < .001$, $\eta^2_p = .33$. The speaker seeing the addressee did not exert a main effect on the gesture rate, $F(1, 31) = 1.41$, $p = .24$. The two factors interacted, $F(1, 31) = 5.34$, $p < .05$, $\eta^2_p = .15$: Seeing the addressee only increased gesture production if the addressee could also see the speaker.

Gesture size Analyses of the average size of representational gestures, shown in Figure 8, revealed a main effect of the speaker being seen by the addressee, such that speakers produced larger gestures when they were seen (*M* = 3.07, *SD* = .46) than when they were not (*M* = 2.53, *SD* = .78), $F(1, 29) = 4.60$, $p < .05$, $\eta^2_p = .14$. The speaker seeing the addressee did not exert a main effect on gesture size, $F < 1$. The two factors did not interact, $F < 1$.

Speech Neither the speaker being seen by the addressee, nor the speaker seeing the addressee exerted a main effect on the total number of words used, the number of filled pauses per 100 words, or the type-token ratio. The speaker being seen by the addressee exerted a main effect on the speech rate, such that speakers spoke faster when they were seen (*M* = 3.03, *SD* = .32) than when they were not (*M* = 2.73 *SD* = .45), $F(1, 32) = 5.07$, $p < .05$, $\eta^2_p = .14$. The speaker

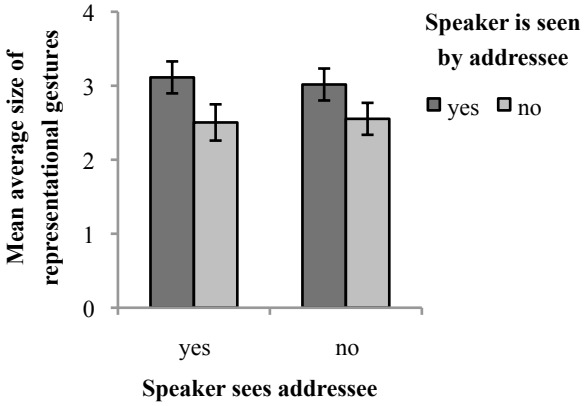

Figure 8: Mean average size of representational gestures, depending on whether the speaker could be seen by the addressee (separate columns) and whether the speaker could see the addressee (x-axis). Bars represent standard errors.

seeing the addressee did not exert a main effect on the speech rate, $F < 1$. The two factors did not interact, $F < 1$.

Other Neither the speaker being seen by the addressee, nor the speaker seeing the addressee exerted a main effect on the duration of the narrations in seconds. There was a significant correlation between participants' speech rate and their gesture rate, $r = .34$, $p < .05$.

Analysis with factors: mutual visibility and mediation

In this analysis we assess the effects of the speaker and addressee being able to see each other and communication being mediated on language production.

Gesture rate Analyses of the number of representational gestures per minute, shown in Figure 9, revealed a main effect of mutual visibility, such that speakers gestured more frequently when interlocutors could see each other ($M = 11.89$, $SD = 7.28$) than when they could not ($M = 3.92$, $SD = 2.30$), $F(1,32) = 19.75$, $p < .001$, $\eta^2_p = .38$. The main effect of mediation showed a trend toward significance, such that speakers gestured more frequently when communication was mediated ($M = 8.75$, $SD = 8.02$) than when it was not ($M = 7.31$, $SD = 5.36$), $F(1, 32) = 3.97$, $p = .06$, $\eta^2_p = .11$. The latter effect was not present without

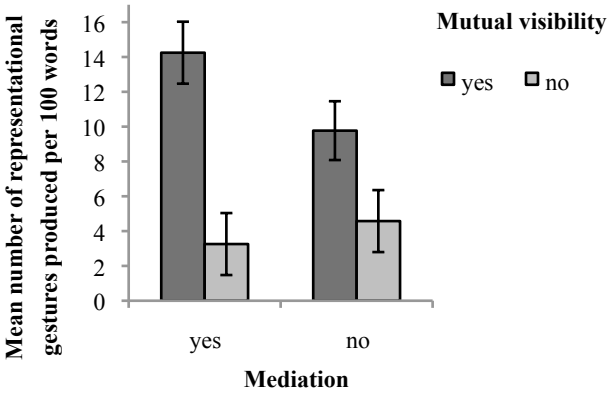

Figure 9: Mean gesture rate depending on whether communication was mediated (x-axis) and whether speaker and addressee could see each other (separate columns). Bars represent standard errors.

speech rate as a covariate ($F < 1$). The two factors did not interact, $F(1, 32) = 1.76$, $p = .19$.

Gesture size Analyses of the average size of representational gestures, shown in Figure 10, revealed a main effect of mutual visibility, such that speakers produced larger gestures when interlocutors could see each other ($M = 3.04$, $SD = .41$) than when they could not ($M = 2.62$, $SD = .52$), $F(1, 31) = 7.64$, $p < .01$, $\eta^2_p = .20$,. Mediation did not exert a main effect on gesture size, $F < 1$, and the two factors did not interact, $F < 1$.

Speech Neither mutual visibility nor mediation exerted a main effect on the total number of words used, or the type-token ratio. Mutual visibility exerted a main effect on the number of filled pauses per 100 words, such that speakers used filled pauses less frequently when interlocutors could see each other ($M = 5.70$, $SD = 3.02$) than when they could not ($M = 7.46$, $SD = 1.93$), $F(1, 33) = 4.18$, $p < .05$, $\eta^2_p = .11$. Mediation did not exert a main effect on the rate of filled pauses, $F < 1$, and the two factors did not interact, $F < 1$. Mediation exerted a main effect on the number of words per second, such that speakers spoke slower when communication was mediated ($M = 2.84$, $SD = .50$) than when it was not ($M = 3.27$, $SD = .51$), $F(1, 33) = 6.64$, $p < .02$, $\eta^2_p = .17$. Mutual visibility did not

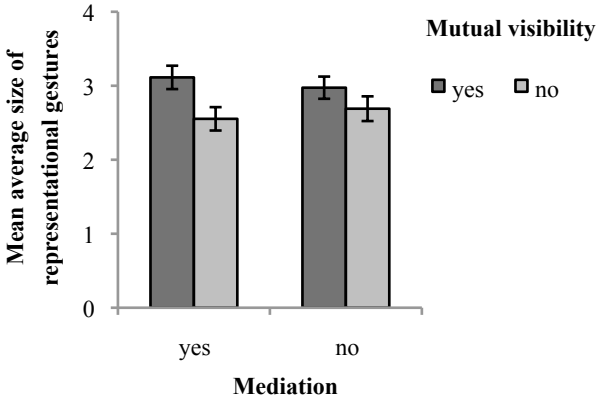

Figure 10: Mean average size of representational gestures, depending on whether communication was mediated (x-axis) and whether speaker and addressee could see each other (separate columns). Bars represent standard errors.

exert a main effect on the speech rate, $F(1, 33) = 1.14$, $p = .29$ and the two factors did not interact, $F < 1$.

Other Neither mutual visibility nor mediation exerted a main effect on the duration of the narrations in seconds. There was a significant correlation between participants' speech rate and their gesture rate, $r = .38$, $p < .05$.

Discussion

Comparison of the Eye-Catcher setting to the setting with mutual visibility through web cams revealed that the extra affordance offered by the Eye-Catcher affected gesture production. Gestures were produced far more frequently when information from gaze could readily be interpreted and speakers could look at their addressee's eyes and see their addressee look at their eyes simultaneously. Therefore, it seems that the decrease in gesture production that we found in our first study when speakers could see their addressee, indeed resulted from the somewhat unnatural gazing behavior of the addressee, which may have caused her to appear inattentive or uninterested. Being able to use gaze may also have affected gesture production more directly, since speakers can use gaze to direct their addressee's attention to their gestures (Gullberg & Kita, 2009).

Provided that interlocutors can see each other (mutual visibility), the Eye-Catcher seems to allow for natural gazing behavior. Therefore, the Eye-Catcher setting is more suitable for addressing how being seen by the addressee and seeing the addressee affect language production. With the Eye-Catcher setting replacing our previous setting with mutual visibility through web cams, the effects of the speaker being visible to the addressee and vice versa become easier to interpret. What we see is that whether the speaker can be seen influences both the frequency and the size of representational gestures, showing that speakers adapt their gesture production to their knowledge of whether the addressee can see them. Seeing the addressee also causes speakers to put more communicative effort into their gesture production, but only if the addressee can see them, and only if the addressee's gaze can readily be interpreted.

Comparison of our mediated settings with mutual or no visibility to similar unmediated settings showed that mutual visibility affected language production similarly, independent of whether communication was mediated or not. This suggests that the Eye-Catcher setting is a close match to face-to-face communication, and that our finding that speakers do take into account their addressee's visual perspective are likely to generalize to unmediated communication. We also found that speakers did not gesture less when communication was mediated, provided that the costs and affordances were a close match to unmediated communication. This supports the grounding framework by Brennan and Lockridge (2006) and is consistent with social presence theory (Short, et al., 1976) as well as social information processing theory (Walther, 1992). Our results do not reveal whether the effect of the lack of interpretability of gaze in web cam settings can be overcome with time, as may be predicted by social information processing theory.

Our global measures related to the content of the narrations showed that the narrations were very similar in all settings. Therefore, it does not seem that the differences in gesture production resulted from differences in speech content. In some of our analyses we again found that being seen by the addressee caused speakers to speak faster, as we found in the analysis of our first experiment. Additionally, one of our analyses showed that speakers produced fewer filled pauses when the addressee could see them. When visual information is unavailable, it may be more necessary to use filled pauses communicatively, indicating that one is still thinking (Clark & Fox Tree, 2002). We also found that speech was faster in unmediated settings, compared to mediated ones. This may

indicate a limited trust in the signal quality in mediated communication. Although we found some correlation between participants' speech rate and their gesture rate, it does not seem that the increased gesture production in some settings resulted from a need to speak faster, because also in unmediated settings speech was faster, without there being an increased gesture rate. We think it more likely that the differences in the communicative settings affected both speech and gesture, with expectedly, manipulations of visibility affecting gesture production more dramatically than speech production.

The differences we found in speakers' gesture production are informative of whether speakers apply their knowledge of the addressee's visual perspective to their language production. Yet do these differences in language production matter to the addressee? Our third study examines whether naive observers are sensitive to some of the differences we found in speakers' language production. It thereby also examines in yet another way how similar mediated and unmediated communication were in our study.

Study 3: Perception study

In this study, we ask participants to rate movie clips from speakers who communicate either through web cams, with the Eye-Catcher, or face-to-face, all with mutual visibility between speaker and addressee. As a measure of how well speakers are perceived to perform the narration task, we ask participants to rate the speakers for their expressivity. If whether communication is mediated influences speakers' behavior most, then we would expect speakers in the Face-to-Face setting to be rated differently from speakers in the two mediated settings. On the other hand, if not mediation but the interpretability of gaze affects language production most, we would expect speakers in the Face-to-Face and Eye-Catcher setting to be rated differently from speakers in the Web Cam setting. If both of these factors play a role, then speakers from each setting may be rated differently. Another possible outcome is that although differences can be found in a formal analysis of gesture, these differences do not matter for how speakers are perceived in terms of their expressivity.

Method

Participants Twenty (17 female) native Dutch first year students from Tilburg University took part in this study. Their mean age was 21, range 18 – 25.

Stimuli We created 27 trials, using 9 movie clips from speakers in each setting with mutual visibility: the Face-to-Face, Eye-Catcher, and Web Cam setting. In addition, we created two practice trials, using data from speakers in an unrelated experiment in which we used a similar cartoon narration task. From each speaker, we chose a fragment of 10 seconds, starting as soon as the speaker started to talk about the third episode of the cartoon. In this episode Sylvester tries to climb up to Tweety's window through an adjacent drainpipe, but gets stopped by a bowling ball, which was thrown into the drainpipe by Tweety. Speakers were visible from the knees up. Each movie clip was preceded by a short beep and an order number that corresponded to a line on the answering form, which was displayed for 2 seconds. After each clip, 4 seconds of blank video were inserted allowing participants time to fill out their answer. After the last movie clip a text was displayed, which indicated that the experiment had ended. The 27 actual clips were presented in a random order. We created two versions, the second one showing the clips in reversed order.

Procedure Participants came to the lab and were asked to rate video-fragments of speakers for how expressive the speaker was. They indicated their answer for each speaker on an answering form, by circling a number on a 1 to 5 scale, '1' meaning 'very little expressive' and '5' meaning 'very expressive'. Participants first saw two practice trials. After the practice trials they were allowed to ask any clarification question on the task, which were answered by the experimenter (without her ever mentioning gesture). The participant then watched the actual fragments, filling out the answering form after each fragment. Half the participants saw the fragments in a certain order and the other half in reversed order. After this task participants filled out a brief questionnaire, which asked for participants' age, native language and what they had based their ratings on.

Statistical analysis We used a Repeated Measures analysis with the setting that the movie clips were taken from as a factor (levels: face-to-face, Eye-Catcher, web cam). For each setting we first computed each participant's mean rating of

the nine movie clips from that setting, which we used as the dependent variable. Pairwise comparisons were done using the LSD method with a significance threshold of .05.

Results

Analyses of participants' ratings of speakers' expressivity revealed a main effect of the setting that the speaker was in, $F(2, 38) = 26.10$, $p < .001$, $\eta^2_p = .58$. Posthoc analysis showed that speakers from the Web Cam setting were rated as less expressive ($M = 2.42$, $SD = .53$) than speakers from the Face-to-Face ($M = 3.07$, $SD = .78$) and Eye-Catcher setting ($M = 3.24$, $SD = .54$). The ratings for speakers from these latter two settings did not differ significantly. The order of presentation of the clips did not exert a main effect on participants' rating, $F < 1$.

In answer to an open question of what participants had based their judgment on, 14 out of 20 participants (70%) spontaneously mentioned that they had partially based their judgments on speakers' hand movements. Participants also mentioned that they had paid attention to the speakers' facial expressions (55%), posture (35%), body language (15%), gaze (10%), eye-brow movements (5%), intonation (40%), use of voice (30%), loudness (10%), clarity of voice (5%), and laughing (15%).

Discussion

Participants were sensitive to the differences in how speakers narrated between the setting with communication through web cams on the one hand, and face-to-face communication and communication through Eye-Catchers on the other. Speakers from the Web Cam setting were rated as less expressive. Most participants took a speaker's hand gestures into account when judging the speaker's expressivity. It therefore seems that producing hand gestures more frequently (as speakers in the Eye-Catcher and face-to-face settings did), is associated with greater expressivity. In addition, there was no perceived difference in expressivity between speakers from the Face-to-Face and from the Eye-Catcher setting, again suggesting that these settings were a closer match than face-to-face interaction and communicating through web cams. Thus, whether communication is mediated or not seems of lesser influence on speakers' language production than whether gaze in mediated settings resembles gaze in unmediated settings.

General discussion and conclusion

When it comes to their gesturing, speakers apply their knowledge of their addressee's visual perspective to their language production. Speakers produced more and larger gestures when they knew their addressee could see them. This suggests that gesturing is at least partly intended communicatively. Other than in previous work, our results cannot be explained by observable changes in the environment of the speaker. Our study therefore supports the interpretations in terms of audience design of earlier studies (Alibali, et al., 2000; Bavelas, et al., 2008; Cohen, 1977; Jacobs & Garnham, 2007; Özyürek, 2002). Speakers are able to adapt their gesturing to their knowledge of their addressee, rather than solely using their own perspective.

This does not prove that speakers never make mistakes when they need to take into account the other interlocutors' visual perspective, as has been found for speech production and interpretation (Keysar, et al., 2003; Wardow Lane, et al., 2006). In our studies, gesture production was used as a global measure of language production. This is different from the use of eye-tracking data as in the study by Keysar et al., which may be able to capture any mistake in language interpretation. It differs similarly from the use of reference production in the study by Wardow Lane et al., which reveals any overspecification. Gesture production thus shows a general trend, rather than capturing every individual mistake.

We found some evidence that gesture frequency is reduced when the addressee seems less interested, as has also been found by Jacobs and Garnham (2007). In addition, we found that gesture size does not seem to be affected by this factor. Thus, although speakers produce fewer gestures when the addressee appears less interested, the gestures they do produce may compare well to gestures directed at addressees appearing more interested. In reality, there was no difference in how interested the addressee was between our conditions, since the addressee always was a confederate. Rather, the addressee was perceived as less interested in one condition, as a result of the costs and affordances offered by the mediated setting. When speakers could see the addressee, but the addressee could not see the speaker, speakers perceived the addressee as less interested. In this case it was not possible for the addressee to show natural gazing behavior. This effect is predicted by the grounding framework (Brennan & Lockridge, 2006), which states that it is not mediation as such that causes interlocutors to

perceive each other differently, but rather the fact that differences in the costs and affordances associated with mediation affect interlocutors' behavior, which in turn leads to different perceptions of each other.

The results we found are consistent with the idea that the more the costs and affordances of mediated and unmediated communication are alike, the more similar language use will be (Brennan & Lockridge, 2006). In our first study, in which we made use of communication through web cams, we saw that seeing the addressee led to a decrease in gesture production rather than an increase. In this case the affordance of seeing the addressee was more similar to face-to-face interaction, but the costs associated with interpreting the addressee's gaze were not. Interpreting the others' gaze is harder when communicating through web cams than when interacting face-to-face. Our second study showed that when this was compensated for by using the Eye-Catcher, seeing the addressee did increase gesture production, such that the gesture rate and size in mediated communication were now comparable to those in face-to-face interaction. The gesture rate was much higher when speaker and addressee could see each other, independent of whether they were in the same room. Therefore, it indeed seems that mediation as such does not have a large effect on language production, as predicted by the grounding framework.

We further showed that observers were sensitive to the differences in speakers' communicative behavior. When participants were asked to judge speakers' expressivity, speakers from a setting with communication through web cams were rated as less expressive than speakers from a face-to-face setting, as well as speakers from a setting in which communication was mediated by Eye-Catchers. This shows that speakers are more expressive when interlocutors can readily interpret each other's gaze. It also confirms that communication through Eye-Catchers resembles face-to-face communication, more so than communication through web cams.

Since narrations were equally long in each setting we used, both in time and in the number of words, and the variation in vocabulary did not differ, it is likely that the communicative setting affected gesture production, rather than the differences in gesture resulting from the narrations being much different. Our studies suggest that speakers may also adjust their speech rate to whether or not their addressee can see them, similar to adjusting their gesture rate and size. In this case too, speakers' knowledge of the addressee seeing them was more important than whether or not they saw their addressee. Speakers spoke faster

when they knew they could be seen, which may indicate they had an intuition that visual information aids speech interpretation. Such an intuition would be consistent with actual findings (e.g. Sumby & Pollack, 1954).

Despite dialogue being possible in the settings we used, the task we used in our studies was mostly a monologue task and the addressee never interrupted the speaker. This has the upside of settings being very similar to each speaker within a certain setting, such that we could get a clear picture of the factors of interest. It also strengthens our case that speakers can use their knowledge of their addressee's perspective, rather than their own direct observation (including the addressee's back-channeling behavior). Yet in future work, it is necessary to look at other factors that come into play when scaling up from monologue to dialogue. Does mediated communication with the Eye-Catcher still pass the test when more turn taking is involved? And what if there are more than two speakers? Our study implies that it is important for interlocutors in such situations to be able to make sense of each other's gaze. Moreover, our studies suggest that when using video-conferencing, it may be important to choose the image such that the hands are visible, since speakers adapt their gesturing to their addressee and thus seem to intend it communicatively.

Acknowledgements

We thank all our participants. We thank the reviewers and editor of the Journal of Computer-Mediated Communication for their comments on the paper this chapter is based on. We thank Nelianne van den Berg, Hanneke Schoormans, Vera Nijveld, and Madelène Munnik for their help in collecting, coding, and transcribing the data. We thank Bernd Hellema for providing a Perl script and Lennard van der Laar for his technical assistance.

References

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language, 44*, 169-188.

Alibali, M. W., Kita, S., & Young, A. J. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes, 15*(6), 593-613.

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. Cambridge: Cambridge University Press.

Bangerter, A., & Chevalley, E. (2007). Pointing and describing in referential communication: When are pointing gestures used to communicate? In I. Van der Sluis, M. Theune, E. Reiter & E. Krahmer (Eds.), *Proceedings of the Workshop on Multimodal Output Generation* (pp. 17-28). Enschede: Universiteit Twente.

Bavelas, J., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse Processes, 15*, 469-489.

Bavelas, J., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language, 58*, 495-520.

Beattie, G., & Shovelton, H. (1999). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of Language and Social Psychology, 18*, 438-462.

Brennan, S. E., & Lockridge, C. B. (2006). Computer-mediated communication: A cognitive science approach. In K. Brown (Ed.), *ELL2, Encyclopedia of Language and Linguistics, 2nd Edition* (pp. 775-780). Oxford, UK: Elsevier Ltd.

Brennan, S. E., & Ohaeri, J. O. (1999). Why do electronic conversations seem less polite? The costs and benefits of hedging. In D. Georgakopoulos, W. Prinz & A. L. Wolf (Eds.), *Proceedings of the International Joint Conference on Work Activities, Coordination, and Collaboration (WACC '99)* (pp. 227-235). San Francisco, CA: ACM.

Brennan, S. E., Xin, C., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition, 106*, 1465-1477.

Cassell, J., McNeill, D., & McCullough, K.-E. (1998). Speech-gesture mismatches: Evidence for one underlying representation of linguistic & nonlinguistic information. *Pragmatics & Cognition, 6*(2), 1-33.

Chawla, P., & Krauss, R. M. (1994). Gesture and speech in spontaneous and rehearsed narratives. *Journal of Experimental Social Psychology, 30*, 580-601.

Chu, M., & Kita, S. (2008). Spontaneous gestures during mental rotation tasks: Insights into the microdevelopment of the motor strategy. *Journal of Experimental Psychology: General, 137*(4), 706-723.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. Levine & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127-149). Washington, DC: APA.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition, 84*(1), 73-111.

Cohen, A. A. (1977). The communicative functions of hand illustrators. *Journal of Communication, 27*(4), 54-63.

Cohen, A. A., & Harison, R. P. (1973). Intentionality in the use of hand illustrators in face-to-face communication situations. *Journal of Personality and Social Psychology, 28*, 276-279.

Cutica, I., & Bucciarelli, M. (2008). The deep versus the shallow: Effects of co-speech gestures in learning from discourse. *Cognitive Science, 32*(5), 921-935.

Effron, D. (1941). *Gesture and environment*. Morningside Heights, NY: King's Crown Presss.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica, 1*, 49-98.

Feyereisen, P., Van de Wiele, M., & Dubois, F. (1988). The meaning of gestures: What can be understood without speech? *Cahiers de Psychologie Cognitive, 8*, 3-25.

Goldin-Meadow, S. (2010). When gesture does and does not promote learning. *Language and Cognition, 2*(1), 1-19.

Goldin-Meadow, S., & Sandhofer, C. M. (1999). Gestures convey substantive information about a child's thoughts to ordinary listeners. *Developmental Science, 2*(1), 67-74.

GreenEyes. (2007). http://www.greeniii.com/index.php

Grice, P. (1989). *Studies in the Way of Words*. Cambridge MA: Harvard University Press.

Gullberg, M., & Kita, S. (2009). Attention to speech-accompanying gestures: eye movements and information uptake. *Journal of Nonverbal Behavior, 33*(4), 251-277.

Hadar, U. (1989). Two types of gesture and their role in speech production. *Journal of Language and Social Psychology, 8*(3-4), 221-228.

Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language, 57*(4), 596-515.

Isaacs, E. A., & Tang, J. C. (1994). What video can and cannot do for collaboration: a case study. *Multimedia Systems, 2*(2), 63-73.

Jacobs, N., & Garnham, A. (2007). The role of conversational hand gestures in a narrative task. *Journal of Memory and Language, 56*(2), 291-303.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica, 26*, 22-63.

Kendon, A. (1988). How gestures can become like words. In F. Potyatos (Ed.), *Crosscultural perspectives in nonverbal communication* (pp. 131-141). Toronto, Canada: Hogrefe.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*(1), 25-41.

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language, 57*(3), 396-414.

Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science, 7*, 54-60.

McNeill, D. (1992). *Hand and Mind: What gestures reveal about thought*. Chicago and London: The University of Chicago Press.

Melinger, A., & Kita, S. (2007). Conceptualisation load triggers gesture production. *Language and Cognitive Processes, 22*(4), 473-500.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2009). The communicative import of gestures: Evidence from a comparative analysis of human-human and human-machine interactions. *Gesture, 9*(1), 97-126.

Özyürek, A. (2002). Do speakers design their cospeech gestures for their addressees? The effects of addressee location on representational gestures. *Journal of Memory and Language, 46*(4), 688-704.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. London: Wiley.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*, 212-215.

Walther, J. B. (1992). Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication Research, 19*(1), 52-90.

Walther, J. B., Slovacek, C. L., & Tidwell, L., C. (2001). Is a Pitcure worth a thousand words? Photographic images in long-term and short-term computer-mediated communication. *Communication Research, 28*(1), 105-134.

Wardow Lane, L., Groisman, M., & Ferreira, V. S. (2006). Don't talk about pink elephants: Speakers' control over leaking private information during language production. *Psychological Science, 17*(4), 273-277.

Chapter 4

Adaptation in gesture: converging hands or converging minds?

Abstract

Interlocutors sometimes repeat each other's co-speech hand gestures. In three experiments, we investigate to what extent the copying of such gestures' form is tied to their meaning in the linguistic context, as well as to interlocutors' representations of this meaning at the conceptual level. We found that gestures were repeated only if they could be interpreted within the meaningful context provided by speech. We also found evidence that the copying of gesture forms is mediated by representations of meaning. That is, representations of meaning are also converging across interlocutors rather than just representations of gesture form. We conclude that the repetition across interlocutors of representational hand gestures may be driven by representations at the conceptual level, as has also been proposed for the repetition of referring expressions across interlocutors (lexical entrainment). That is, adaptation in gesture resembles adaptation in speech, rather than it being an instance of automated motor-mimicry.

This chapter is based on:

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (In Press). Adaptation in gesture: Converging hands or converging minds? *Journal of Memory and Language.*

Introduction

Suppose Mary and John are discussing a route through a city. Then if Mary refers to an alley as 'the narrow street', it is likely that John will also use this expression when subsequently referring to that alley, rather than using a completely different expression, such as 'the alleyway'. In addition to saying 'the narrow street', Mary may hold up her hand in front of her, with her fingers pointing toward the end of the alley, thereby indicating the direction that the alley runs in. This hand gesture may subsequently be repeated by John, when he talks about the alley again. Would such a repetition of a gesture be similar to a repetition of a referring expression?

When people interact in dialogue, they adapt to each other in many ways (for an overview, see Branigan, Pickering, Pearson, & McLean, 2010). For example, Brennan and Clark (1996) showed that interlocutors tend to repeat each other's referring expressions, a process known as *lexical entrainment*. Apart from verbal adaptation, interlocutors can also repeat each other's nonverbal behaviors (e.g. Chartrand & Bargh, 1999), among which are the hand gestures that many people produce spontaneously while talking. It has been found that people indeed repeat such hand gestures of each other (De Fornel, 1992; Holler & Wilkin, 2011; Kimbara, 2008; Tabensky, 2001). Yet how similar are these repetitions in gesture to repetitions in speech? Do similar processes underlie adaptation in both speech and gesture? And what is the role of meaning? Do interlocutors produce similar hand gestures because they construct similar representations of meaning, or are they merely copying each other's movements?

Mimicry and adaptation

Mimicry and adaptation in interaction have been studied extensively. Chartrand and Bargh (1999) for instance, found that participants were more likely to shake their foot during a conversation if their confederate conversation partner did so as well, and similarly for rubbing one's face with one's hand. According to Chartrand and Bargh (1999), although such mimicry may act as a kind of 'social glue', intent or conscious effort are not required for it to occur. They state that "the mere perception of another's behavior automatically increases the likelihood of engaging in that behavior oneself", p. 893. This is known as the *perception-behavior link.*

Pickering and Garrod (2004) propose that similar automated priming may frequently underlie the repetition of linguistic forms across interlocutors, which they call *alignment*. They propose that at each linguistic level, "[t]he activation of a representation in one interlocutor leads to the activation of the matching representation in the other interlocutor directly", p. 177. For example, after perceiving a sentence in the passive voice, people are more likely to produce a passive voice as well (Bock, 1986). In Pickering and Garrod's *interactive alignment account*, this is explained as the speaker's representation of the passive voice being more activated as a result of perception, and therefore being a more likely candidate for subsequent production. Thus, it is assumed that representations are shared between comprehension and production (parity of representation).

In the interactive alignment account, the repetition across interlocutors of a linguistic form at any one level (e.g. the syntactic level) can happen without representations at other linguistic levels playing a critical role. For example, repetition of the same syntactic structure could happen when a similar, but also when a very different lexical form or meaning is being expressed than was perceived. Although stronger effects have repeatedly been found when the same word was repeated, syntactic alignment indeed also occurred when different words (with different meanings) were produced than were perceived (Cleland & Pickering, 2003; Pickering & Branigan, 1998).

It thus seems that adaptation of syntactic structures can occur independently of representations at the lexical and the conceptual level. Yet it is sometimes hard to see how adaptation at the lexical level can occur without the conceptual level being involved. It is rare that speakers choose the same referring expression as their interlocutors, whereas they do not want to express the same meaning. Therefore, when looking at the repetition of forms that carry propositional meaning, we need to describe how meaning is involved in the repetition of form. In Pickering and Garrod's view, even though alignment of form tends to be linked to alignment of meaning, it does not have to be. One way their model could account for such a link is that it also assumes connections between the representations at different linguistic levels within a speaker. This means that activation of a representation at the lexical level could lead to activation of a representation at the conceptual level, and vice versa. This way, when during perception the representation for the word form 'alley' is associated with a certain representation of meaning, the connection between those two

representations is strengthened. When the same representation of meaning is subsequently activated in the production process, the representation of the word form for 'alley' receives activation through this connection, making the word 'alley' a more likely candidate for production. Although the model can thus account for the link between alignment of meaning and form, its unique contribution lies in that it is also possible for alignment of form to occur without the alignment of meaning, as alignment at each linguistic level can happen independently of other levels.

Brennan and Clark (1996) do propose that when interlocutors use the same words to refer to the same objects, this is because they use similar conceptualizations of those objects. For example, suppose a particular object could be thought of as a document, a picture, or a map. When a speaker refers to it as a 'map', she conceptualizes the object for the current purpose as such. If the addressee agrees with this conceptualization and a *conceptual pact* is formed, both interlocutors can subsequently use the utterance 'map' as a reference to both the object and the particular conceptualization of it. Thus, for both interlocutors a certain utterance is linked to a certain object, as well as to a certain representation of meaning at the conceptual level. In this view, representations of meaning are necessarily involved in the repetition of referring expressions across interlocutors, and for interlocutors to use similar expressions they need to have similar representations at the conceptual level. Before exploring what answers current models on gesture production and earlier studies on adaptation in gesture can provide to the question of whether or not the conceptual level is necessarily involved in the repetition of hand gesture forms across interlocutors, we first specify what we mean by gestures.

Co-speech gestures

When talking, many people move their hands and arms around without the objective of directly manipulating their environment. Rather, such movements seem to be part of their communicative effort (Kendon, 2004). For example, raising one's arm while extending one's index finger toward an object could disambiguate the question "Can you hand me that?". Kendon (1988) recognizes different kinds of hand gestures, which McNeill (1992) put on a continuum of how conventional and language-like the hand gestures are. On one end of this continuum are sign languages, in which signs have a conventional meaning. At the far other end is *gesticulation.* This is the production along with speech of

gestures that are not embedded in the structure of speech. For example, moving the hands along the upper body as though running while saying: "He went away". Gestures at this end of the continuum are the most idiosyncratic and their interpretation is highly dependent on the accompanying speech (Feyereisen, Van de Wiele, & Dubois, 1988).

Gestures that fall into Kendon's category of gesticulation are generally divided into several subcategories (e.g. McNeill, 1992). One broad distinction can be made based on whether a gesture depicts some of the content of the message a speaker is trying to convey, or whether it rather structures the conversation (e.g. Bavelas, Chovil, Lawrie, & Wade, 1992), or emphasizes certain parts of speech (e.g. Effron, 1941; Ekman & Friesen, 1969; Krahmer & Swerts, 2007). In this paper we focus on gestures that express some of the content a speaker is conveying, which are known as *illustrators* (Ekman & Friesen, 1969), or *representational gestures* (McNeill, 1992). Elements of such a gesture's physical form, like the shape and orientation of the hand, the direction and size of the movement, and where it is performed relative to the speaker (Müller, 1998), can be repeated in subsequent gestures by the same or another speaker. Yet importantly, since these gestures are among the least conventional on Kendon's continuum, there are many different ways in which the same content could potentially be expressed in co-speech gesturing.

Gesture and speech production and perception

Gesture and speech have been found to be linked temporally (Chui, 2005), structurally (Kita, et al., 2007), pragmatically (Enfield, Kita, & De Ruiter, 2007), and semantically (McNeill, 1992). Therefore, gesture and speech production are somehow coordinated. McNeill's Growth Point theory states that speech and gesture co-express idea units, which develop themselves into utterances (McNeill, 1992, 2005). That is, speech and gesture co-develop over time, into an utterance. Therefore, it is no surprise that current frameworks of gesture production are based on a framework of speech production, specifically the framework of Levelt (1989).

Levelt's *blueprint for the speaker* discriminates between a conceptualization, a formulation, and an articulation stage. Based on this model, De Ruiter (1998, 2000, 2007) proposes that a communicative intention is formed in the conceptualization stage, which is then passed on to two parallel formulation stages: one for gesture and one for speech. Each of these formulation stages

leads to its own articulation stage, either the articulation of gesture or the articulation of speech. This architecture is called the *postcard architecture*, because rather than assuming that gesture directly reflects thought, it assumes that a communicative intention underlies gesture production. Therefore, the metaphor of postcards from the mind may be more accurate than gesture being a window into the mind (McNeill, 1992). In a postcard model, gesture and speech production share the stages up until the formulation of a communicative intention, and then continue in parallel, but separately.

This is different from models based on the *interface architecture* proposed by Kita and Özyürek (2003). In interface models, the processes of gesture and speech formulation interact with each other online. In the interface model they propose, the message to be communicated is not fully determined by one conceptualization module, but also by two separate generators: the action generator for gesture and the message generator for speech. The action generator has access to spatial and motoric components in working memory as well as to a model of the environment, while the message generator has access to propositional components in working memory and the discourse model. Importantly, there are bidirectional links between the action and message generator, as well as between the message generator and the speech formulator. This means that constraints on how a message can be expressed in speech, can also influence how it is expressed in gesture. Thus, the content of gesture is not fully specified by the communicative intention alone, but also by the features of imagined or real space, and online feedback from the speech formulator via the message generator.

Both the postcard and the interface architecture focus on the production of speech and gesture, and thus model the speaker. It is not specified what representations are shared between production and comprehension or where the links between these processes are. When it comes to adaptation in gesture, both gesture production and perception are involved. Levelt assumes that lemmas and forms are shared between speech formulation and speech comprehension within a speaker. This assumption is also incorporated in the interactive alignment account proposed by Pickering and Garrod (2004). Their model includes multiple interlocutors, allowing it to explain adaptation of one interlocutor to another. It is however non-specific about gesture.

We can apply the interactive alignment account to adaptation in gesture in two ways. Firstly, if we assume that gesture forms are represented at their own

linguistic level, that they are shared between perception and production, and that interlocutors can align to each other's forms directly, adaptation in gesture could occur without the conceptual level playing a critical role. This would be a low-level account of adaptation in gesture. Secondly, if we assume that representations at the conceptual level are shared between speech and gesture comprehension and production, and that there are bidirectional links between the different linguistic levels within a speaker, then adaptation in gesture could also occur via the conceptual level. A conceptual representation that is activated as a result of the perception of speech and gesture can subsequently influence the production of speech and gesture.

A postcard model can also provide such a higher-level account: A communicative intention that is activated through the perception of gesture and speech, could subsequently inform gesture and speech production. This route is also possible in Kita and Özyürek's interface model, but this model would still allow for speech and gesture to be coordinated during the formulation stage of production as well. Additionally, the interface model may be able to account more readily for adaptation of gesture forms at lower levels than the conceptual level, since the action generator has access to information specifically relevant to gesture, and can coordinate the generation of a gesture with the generation of a spoken message directly. Since neither of these models includes another interlocutor or comprehension of speech and gesture explicitly, these accounts remain speculative.

It seems that current models of gesture and speech production, combined with the interactive alignment account allow for both a low-level explanation (not involving representations of meaning) and a higher-level explanation (involving representations of meaning) of adaptation in gesture, depending on what representations are shared between perception and production, and at what levels these processes are linked. Can studies on adaptation in gesture tell us more about whether representations of meaning are necessarily involved when a perceived gesture form is subsequently produced?

Repetitions of co-speech gestures across interlocutors

Bergman and Kopp (2009) found that properties of a shape to be described influence what representation technique is chosen in gesture. This means that if two speakers are discussing the same objects, their gestures may look similar as a direct result of the content they are expressing, rather than because they adapt

their representations of meaning to each other, or because they mimic each other's movement. So the first question to be answered is whether adaptation occurs in gesture at all.

Compared to adaptation in speech, relatively few studies (e.g. De Fornel, 1992; Holler & Wilkin, 2011; Kimbara, 2006, 2008; Parrill & Kimbara, 2006; Tabensky, 2001) have addressed adaptation in co-speech gesturing. Kimbara (2008), for example, studied dyads while they were jointly retelling an animated cartoon to a camera, narrating such that a third person could understand. She found that when the two speakers could see each other, their representational gestures looked more similar than when they were separated by an opaque screen. This shows that adaptation occurs in gesture: speakers adapted the form of their gestures to the form of another speaker's gestures. This is an important finding. It shows that similarities in interlocutors' gestures did not arise solely because their production tasks were similar, but that seeing each other was critical. However, this study does not reveal whether interlocutors adapted to each other's gestures due to automated motor-mimicry following the perception-behavior link, without intervention of the conceptual level, or whether certain gesture forms were linked to certain representations of meaning at the conceptual level, which caused the forms to be repeated when the same concepts were discussed.

Parill and Kimbara (2006) found that gestures can also be repeated by an observer to a conversation, while subsequently addressing yet another person. In this study, participants were asked to watch a stimulus movie in which two women were discussing what route to take through a model city in front of them. They found that when participants watched a movie in which the women repeated more features of each other's hand gestures, they were more likely to produce these features in their own gesturing later on, while retelling the stimulus movie to the experimenter, compared to when they had seen a movie in which the women repeated fewer of each other's gesture features. A similar yet independent effect was found for verbal repetitions. Parill and Kimbara conclude that people are very sensitive to repetitions across interlocutors in both gesture and speech. This study also shows that the repetition of perceived gesture forms does not happen exclusively between conversation partners, but also when addressing a different person than the one who produced the original gesture. This suggests that adaptation of gesture forms in communication is not always part of an implicit negotiation process on meaning (Brennan & Clark, 1996).

However, the study does not reveal whether participants repeated the observed gesture forms simply because they have a tendency to repeat observed behaviors (Chartrand & Bargh, 1999), or whether they repeated the relations between gestures, objects and representations of meaning used by the people they had observed.

Some indication that the relation between a representation of gesture form and a representation of meaning may be involved in adaptation in gesture forms comes from a study by Tabensky (2001). She observed spontaneous conversation between two participants who were freely discussing a certain topic, and found that in analogy to how verbal information can be repeated literally or be paraphrased, the same information from gesture could be repeated by another interlocutor with either a similar or a very different gesture. This tentatively suggests that interlocutors' representations were converging at the conceptual level, rather than at the level of gesture forms. Also, information contained in one interlocutor's verbal description was found to end up in the other interlocutor's gestures, and vice versa. This suggests that there are links between representations of speech forms and gesture forms, possibly through the conceptual level. Tabensky found that gesture rephrasing only occurred at places where speakers were creating their own meaningful expressions, and not when literally repeating the other person verbally, in which case no gestures were produced. She therefore concludes that gesturing may be intrinsically related to the creation of meaning. This goes well with the idea that similarities in peoples' gestures result from similarities at the conceptual level, where communicative intentions are formed. However, in addition to these observational results, more empirical evidence is needed to support this causal claim.

Cassell, McNeill, and McCullough (1998) studied the relation between representations of meaning and gesture experimentally. They propose that both the perception of gesture and the perception of speech contribute to the construction of an internal representation. This representation of meaning can in turn inform gesture and speech production. This theory is based on an analysis of speakers who retold a story that they had seen a speaker tell in a movie clip. The speaker in the stimulus movie sometimes conveyed different information in gesture than he did in speech. For example, he would say "lure" and gesture either a grabbing or a beckoning motion. The information from the speaker's gestures was found to affect both participants' gestures and their speech. Thus, it seems that information obtained from gesture can subsequently be expressed in

speech and in gesture, which suggests a representation of meaning being shared between gesture and speech interpretation and production. Yet are such representations key to the repetition of gesture forms across interlocutors?

Holler and Wilkin (2011) propose that reproducing each other's gestures in dialogue contributes to the creation of mutually shared understanding. For example, copying a gesture could signal the acceptance of an accompanying referring expression. This would mean that the reproduction of a gesture form signals that a similar representation of meaning has been created at the conceptual level. However, in this study the definition of a copied gesture included that the gesture had the same meaning as the original gesture. Therefore, this study gives a functional account of the copying of gestures assuming that meaning is involved, rather than questioning whether representations of meaning are necessarily involved in the reproduction of gesture forms across interlocutors.

In sum

On the one hand we see that interlocutors adapt to each other's gesture forms less when interlocutors cannot see each other, and that people do not exclusively adapt their gestures to their conversation partner. This goes well with a model in which perceiving a certain gesture form activates a representation of that gesture form, which is subsequently more likely to be selected for production, analogous to Pickering and Garrod (2004). Adaptation in gesture would then be driven by representations of form converging across interlocutors, rather than representations of meaning.

On the other hand, we see that information from one interlocutor's speech can subsequently be expressed in another interlocutor's gesturing, and also that information from one interlocutor's gestures can subsequently be expressed verbally by another interlocutor. This suggests that the same representations of meaning may underlie the production of both speech and gesture, and that similarities in speech and gesture forms across interlocutors may result from them having constructed similar representations of meaning. This is consistent with the models of speech and gesture production proposed by De Ruiter (2000) and Kita and Özyürek (2003). But is this really the case? To what extent is the gesture form produced by one interlocutor determined by the mere perception of a gesture form produced by another interlocutor, and to what extent is it determined by the producer's representation of meaning at the conceptual level?

Present study

We use an experimental approach to address this question. First, we seek to confirm whether seeing a certain representational gesture while hearing certain content in speech, increases the likelihood of producing that same gesture later on, while expressing the same content. For this a speaker in a stimulus movie either does or does not perform certain gestures. These gestures were chosen such that they added very little meaning to the verbal description, for example the speaker moved his arms as though running while talking about running (as opposed to for example making this gesture while only mentioning 'going', in which case it would add much more information to the verbal description). This way, we can observe whether any similarity between the originally perceived gesture and a subsequently produced gesture results from expressing similar content, or whether it is necessary for the producer to actually observe the original gesture.

Second, we address the question of whether hand gestures being meaningful is relevant for their repetition across speakers to occur. We do so by keeping the gesture forms constant across conditions, but varying whether a form matches the content of the co-occurring speech. For example, the speaker would produce the above-described running gesture while talking about looking through binoculars. We predict that if meaning is involved in the repetition of gesture forms, participants will repeat only those gestures whose form matches the content of the concurrent speech. Contrastingly, if the property of carrying meaning is not relevant for the copying of form to occur, or in other words, the conceptual level is not involved, gesture forms will be repeated equally often, independent of where in the narration they occur. Together, these first two experiments address the issue of whether hand gestures are similar to lexical forms when it comes to their repetition across interlocutors, or whether they are better compared to behavioral mimicry, such as the mimicking of each other's foot shaking and the rubbing of one's face.

We then zoom in on the role of representations of meaning in the repetition of gesture forms across interlocutors. We test whether a perceived gesture form influences the construction of a representation of meaning, which subsequently influences gesture production. We do so by looking at different physical features of a gesture. Suppose that certain features of a perceived gesture form give rise to the construction of meaning. Then when this meaning is subsequently expressed in gesture, all features of the produced gesture will be consistent with

this meaning. So we would expect that features of the perceived gesture that were inconsistent with the meaning constructed would not be repeated. On the other hand, if the repetition of gesture forms is not mediated by a representation of meaning, each combination of perceived features could subsequently be produced. This means that we would expect a literal repetition of the perceived gesture, or any of its features, rather than all features being consistent with a certain meaning.

Experiment 1A: Repetition of gesture form

In this experiment we test whether perceiving certain gestures while hearing a story, increases the chance of performing those gestures later on, while retelling the same story.

Method

Participants Participants to this and the following experiments were all adult native speakers of Dutch and they only took part in one of the experiments. Most of them were students at Tilburg University. All of them gave informed written consent for the use of their data. Experiment 1A had 38 (28 female) participants.

Stimuli Two movie clips were created, in which the same male speaker told the same story of an animated cartoon ('Canary Row' by Warner Brothers) as though he had just watched it. Each movie clip consisted of ten fragments. It started with a short introduction in which the speaker stated that the cartoon was a Tweety and Sylvester movie in which Sylvester (a cat) tries to capture Tweety (a pet bird). Then followed eight fragments in each of which the speaker described one episode of the cartoon, which corresponds to one attempt of Sylvester to catch Tweety. These fragments lasted about 15 seconds each. The final fragment consisted of a short closure. Blank video was inserted in between the fragments, allowing for the movie to be paused at appropriate times. The speaker was seated in a chair and looked straight into the camera. The image showed the upper-body of the speaker in front of a white wall, see Figure 1.

Figure 1: Example of the repetition of a target gesture (left) by a participant (right).

The two versions of the stimulus movie differed only in the number of representational hand gestures the speaker produced. In one version, he produced a representational gesture depicting an action for each episode of the cartoon. These gestures were based on retellings of participants in a previous study (Mol, Krahmer, Maes, & Swerts, 2009). They consisted of:

Binoculars: Two hands (cylinder shaped) are held in front of the eyes as the speaker looks through them, representing looking through binoculars. The hands are moved slightly toward the face and back, such that the fingers describe the cylindrical shape of the binocular tubes.

Drainpipe: Two hands/arms make climbing/grabbing motions while moving upward, depicting climbing up the drainpipe.

Rolling ball: Two hands spin around each other from the wrists, while held in front of the speaker, representing rolling.

Money tin: Right hand imitates the holding and shaking of a money tin.

Creeping: Hands (flat, palms down) and arms are moved forward one by one, imitating a creeping motion.

Throwing the weight: Two hands (fingers spread, palms facing each other) are held about 30 cm apart, while a motion is made starting at the head and moving forward in an arc, as though throwing something big away from oneself.

Swinging: Two hands are held on top of each other and quickly make a grabbing motion above and to the side of the speaker's head, representing the grabbing of a rope.

Running: Arms are moved as while running, close to the body of the speaker.

The verbal descriptions of these events were rich, such that the additional information expressed in gesture was minimal. In the other version of the movie clip no representational gestures were produced. No other hand gestures were produced in any of the two versions and care was taken to make the verbal descriptions, body posture, facial expressions, intonation, voice quality, and other prosodic factors maximally similar across the two versions.

In both versions, the speaker used eight target phrases. These were unusual wordings, for example "as a full-blown Tarzan" (Dutch: als een volleerde Tarzan) or "the yearly spring call of the canary" (Dutch: de jaarlijkse lenteroep van de kanarie). These target phrases were the same in both versions. Inclusion of these target phrases allows for comparison of adaptation in gesture and speech and serves as a control measure.

Procedure Participants came to the lab and were assigned randomly to the 'Gestures' or 'No Gestures' condition. They read the instructions, which explained the task as a memory task in which they had to watch video fragments of a speaker telling a story and were asked to subsequently retell these story fragments to the experimenter. Participants were instructed to take as much time as needed when retelling the stories. They were given the opportunity to ask further clarification and once all was clear the experiment started.

Participants first watched the introductory fragment, which they did not have to retell. Then they watched the fragments describing the cartoon episodes, one at a time. After each fragment, participants paused the movie and turned ninety degrees such that they were facing the experimenter while they retold the story. The experimenter was blind to the experimental condition. A camera was placed to the side of the experimenter, recording the participant. Participants were told they were videotaped in order to facilitate our analyses afterward. The experimenter did not interrupt the participants and did not produce any hand gestures, but did show other nonverbal signs of listening to their story in a

natural way (such as by eye-gazing behavior and head movements). Finally, participants watched the last fragment, which they did not have to retell. Note that participants only saw one of the two stimulus movies of a speaker retelling the original cartoon movie and did not see the animated cartoon themselves. The entire experiment took place within twenty minutes.

Coding Each gesture in the stimulus movie of the Gestures condition occurred with a given content unit in the verbal narration. We coded participants' representational gestures produced with those content units in their own narration. We refer to these points in the narration as *target moments*. For example, the binoculars gesture in the stimulus movie was produced while the speaker said that Sylvester was looking at Tweety through binoculars. In this case we looked at participants' gestures while they were describing the event that Sylvester looked at Tweety (the target moment). Gestures from each condition that matched the corresponding gesture in the stimulus movie of the Gestures condition in the hand shape used, the location and movement of the hands, and the event expressed in the concurrent speech, were labeled as *target gesture*. For an example, see Figure 1. If a different gesture was produced with the content unit of the original target gesture this was labeled as a *different gesture*, and if no gesture was produced this was labeled as *no gesture*.

Initially, all gestures were coded by a single coder. Reliability was assessed by having a second coder code a random sample of 20% of the retold fragments, N = 60. The two coders agreed on 88% of the labels. The inter coder reliability for the raters was Cohen's Kappa = .69, indicating substantial agreement (Landis & Koch, 1977). Given the observed marginal frequencies of the labels, the maximum value of Kappa was .79. In our analyses, we used the coding of the first coder.

If a full target phrase was used by participants this was labeled as a *verbal repetition*. If participants repeated one or more (yet not all) content words of the target phrase this was labeled as *partial verbal repetition*. A (partial) verbal repetition was counted as such regardless of when in the participants' retelling it occurred, yet unsurprisingly they occurred only during retellings of the matching episode in the stimulus movie.

Statistical analysis We compared the means across conditions for all dependent variables in this and the following experiment. When Levene's test for equality

of variances was significant, we used the unequal variance t-test. We report mean differences between the compared conditions (M_D), 95% confidence intervals (CI) and we report ω^2 as a measure of effect size.

Results

Gesture The number of *target gestures* produced at target moments was higher in the Gestures condition (M = 1.28, SD = 1.84) than in the No Gestures condition (M = .11, SD = .32) (M_D = 1.17, 95% CI = .24, 2.09), t(18.05) = 2.65, $p < .02$, ω^2 = .14. We did not find an effect of condition on the number of *different gestures* produced at target moments (Gestures: M = .94, SD = 1.11, No Gestures: M = 1.00, SD = 1.65), t(34) = .12, p = .91. There was a trend toward significance for the number of target moments at which *no gesture* was produced, which tended to be higher in the No Gestures condition (M = 6.89, SD = 1.64) than in the Gestures condition (M = 5.78, SD = 2.10) (M_D = 1.17, 95% CI = .24, 2.09), t(34) = 1.77, p = .09, ω^2 = .06.

Speech We did not find a significant difference in the number of verbal repetitions between the Gestures (M = 1.61, SD = .99) and No Gestures condition (M = 1.50, SD = .92) (M_D = .11, 95% CI = -.53, .76), t(34) = .35, p= .73, nor in the number of partial verbal repetitions across the two conditions (Gestures: M = 1.83, SD = 1.34, No Gestures: M = 1.56, SD = 1.38) (M_D = .45, 95% CI = -1.20, .64), t(34) = .61, p = .54.

Discussion

Participants produced certain representational gestures more often if they had seen these gestures in the stimulus movie. Like Kimbara (2008), we found that expressing the same content was not sufficient for these repetitions to occur. Neither was seeing the speaker or an addressee, which participants did in both of our conditions. Rather, seeing the target gestures performed by the speaker in the stimulus movie increased the likelihood of participants producing the same gestures during their own narration to a different addressee. The facts that participants repeated some target phrases and that there was no difference between the two conditions in the number of verbal repetitions show that participants did adapt to the speaker, regardless of whether he gestured.

The fact that participants reproduced gesture forms even though the addressee was different from the speaker in the stimulus movies, suggests that low-level processes, such as priming, may underlie these repetitions, rather than the construction of shared meaning across interlocutors. Seeing a certain form increased the likelihood of producing that form later on. Yet we do not know to what extent a representation of meaning was involved as well. How important was it that these gestures carried meaning for the repetition of their form to occur?

Experiment 1B: Repetition of gesture form and the semantic context

In this experiment we test whether the repetition of a gesture's form across speakers depends on the gesture's meaning in relation to the meaning of the concurrent speech.

Method

Participants Forty-seven participants (33 female) volunteered for this study.

Stimuli Again two stimulus movies were produced, which were similar to the clips in the previous experiment. The first stimulus movie was made in the same way as the one containing representational gestures in the previous experiment (1A). In the second stimulus movie, the speaker produced one representational gesture per episode as well, however this time the gesture did not match the speaker's verbal description. A gesture from another episode was produced instead of the original gesture, along with the original content unit in the verbal description. For example, instead of the 'binoculars' gesture, the speaker produced the 'running' gesture while verbally referring to the event involving binoculars, see Figure 2. The same speaker and the same target phrases were used as in the previous experiment and again care was taken to make the verbal descriptions, body posture, facial expressions, intonation, voice quality, and other prosodic factors maximally similar.

Procedure The procedure was the same as in the previous experiment. In the Congruent condition, 24 participants saw and retold the stimulus movie in which

Figure 2: Congruent (left) and Incongruent (right) gesture for the content unit 'Sylvester looks through binoculars'.

the gestures matched the content of the concurrent speech. In the Incongruent condition, 23 participants saw and retold the stimulus movie in which the gesture forms were mixed up and did not match the content of the concurrent speech. When asked by the experimenter, none of the participants showed any indication of suspecting that the experiment was about the repetition of hand gestures.

Coding In the stimulus movies, gestures occurred at a certain content unit in the verbal narration. Initially, we coded participants' representational gestures produced with those content units in their own narration, that is, at the *target moments*. We coded gestures that matched the corresponding gesture in the movie that the participant had seen in the hand shape used, the location of the gesture and the movement involved in the gesture as *target gesture*, similar as before. We added the label *partial target gesture*, for gestures that matched the gesture in the stimulus movie in two out of these three features (hand shape, location, movement). This time, it was also possible that a participant spontaneously produced the target gesture shown in the other condition. For example, if a participant was shown the running gesture with a description of the event in which Sylvester looks at Tweety through binoculars, this participant could still produce the binoculars gesture while narrating that Sylvester looked at Tweety. Such cases were labeled as *target gesture other condition*.

If a different gesture was produced at a target moment this was labeled as *different gesture*, and if no gesture was produced this was labeled as *no gesture*. If one thinks of gesture and speech as fully separate behaviors, a theory of motor-mimicry is non-specific about the moment at which a target gesture will be reproduced. Therefore, we also looked for target gestures from the moment a

gesture was presented to a participant till the end of the experiment, rather than at target moments only. Verbal repetitions were coded in the same way as before.

Initially all gestures were coded by a single coder. Reliability was assessed by having a second coder code a random sample of 20% of the retold fragments, N = 75. The two coders agreed on 91% of the labels, Cohen's Kappa = .82, indicating almost perfect agreement (Landis & Koch, 1977). Given the observed marginal frequencies of the labels, the maximum value of Kappa was .87. In our analyses, we used the coding of the first coder

Results

Gesture The number of *target gestures* produced at target moments was higher in the Congruent condition (M = .79, SD = 1.10), than in the Incongruent condition (M = .04, SD = .21) (M_D = .75, 95% CI = .28, 1.22), $t(24.71) = 3.26$, $p < .01$, $\omega^2 = .17$. This was also the case for target gestures that were produced at any time from the presentation of the gesture till the end of the experiment (Congruent: M = .79, SD = 1.10, Incongruent: M = .09, SD = .29) (M_D = .71, 95% CI = .23, 1.18), $t(26.26) = 3.03$, $p < .01$, $\omega^2 = .15$. We found no effect on the number of *partial target gestures* produced at target moments (Congruent: M = .96, SD = 1.23, Incongruent: M = .57, SD = .73), $t(45) = 1.32$, $p = .19$.

At target moments, participants in the Congruent condition never produced target gestures from the Incongruent condition. Yet participants from the Incongruent condition sometimes produced target gestures from the Congruent condition at the target moment of those gestures. Thus, the number of *target gestures from the other condition* was higher in the Incongruent (M = .17, SD = .39) than in the Congruent condition (M = .00, SD = .00) (M_D = 1.74, 95% CI = .01, .33), $t(22) = 2.15$, $p < .05$, $\omega^2 = .07$. These included both gestures that had been presented to participants from the Incongruent condition with earlier episodes in their stimulus movie and gestures that these participants had not yet seen. Thus, as in the previous study, participants sometimes spontaneously produced target gestures that they had not seen. Therefore, we compare the number of target gestures from the Congruent condition that were spontaneously produced in the Incongruent condition (*target gesture other condition)* to those produced in the Congruent condition (*target gesture)*. The number of target gestures from the Congruent condition produced at their target moments was

lower in the Incongruent condition (M = .17, SD = .44) than in the Congruent condition (M = .79, SD = 1.10) (M_D = .62, 95% CI = .13, 1.11), $t(45) = 2.58$, $p <$ .02, ω^2 = .11. See Figure 3 for an overview of these results.

Participants in the Incongruent condition more often produced *no gesture* at the target moments (M = 6.87, SD = 1.10) than participants in the Congruent condition (M = 6.00, SD = 1.38) (M_D = .87, 95% CI = .04, .14), $t(45) = 2.38$, $p <$ .05, ω^2 = .09. We did not find an effect on the number of *different gestures* produced at target moments (Congruent: M = .35, SD = 1.15, Incongruent: M =.25, SD = 1.45), $t(45) = 2.38$, p = .80.

Speech Partial verbal repetitions occurred more often in the Congruent condition (M = 1.92, SD = 1.06), than in the Incongruent condition (M = 1.26, SD = .92) (M_D = .66, 95% CI = .07, 1.24), $t(45) = 2.27$, $p < .05$, ω^2 = .08. We did not find a significant effect for full repetitions across the two conditions (Congruent: M = 1.13, SD = .34, Incongruent: M = 1.09, SD = .60) (M_D = .04, 95% CI = -.25, .32), $t(45) = .27$, p = .79.

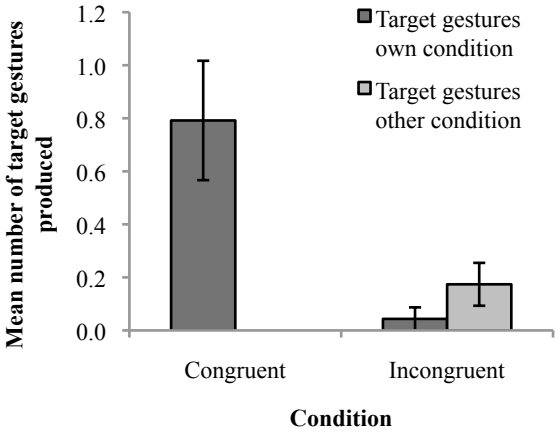

Figure 3: Mean number of target gestures from the participant's own and the other condition (re)produced by participants in the Congruent and Incongruent condition. Bars represent standard errors.

Discussion

Participants repeated gesture forms more frequently if they had been presented to them in a linguistic context in which they were meaningful. Only one target gesture was repeated in the Incongruent condition. In this case the verbal retelling was adjusted such that it matched the gesture. Instead of telling that Sylvester climbed up the drainpipe as had been told in the stimulus movie, the participant said that Sylvester moved past Tweety while producing the gesture, which was a reproduction of the gesture depicting the swinging event. (This was still counted as a repetition of the target gesture at a target moment, since it occurred with the content unit of Sylvester's movement toward Tweety.) These results suggest that representations of meaning do play a role in the repetition of meaningful gestures across speakers. If observing a form would lead to the repetition of that form directly and automatically, or if seeing hand movements alone would cause participants to produce more target gestures, then there would be no reason why gestures that were not meaningful in the linguistic context were less likely to be repeated. This sets the repetition of representational hand gestures across interlocutors aside from the mimicking of behaviors that do not carry propositional meaning, such as the shaking of one's foot or other hand movements.

However, there is a possible confound. It may have been the case that participants were just less likely to adapt to a speaker who came across as somewhat incoherent, due to his non-matching gestures. The result that participants also repeated the target phrases a bit less when they saw the non-matching gestures is consistent with this explanation. In experiment 2 we examine the relation between the repetition of gesture forms and representations of meaning in more detail, this time with a more subtle manipulation of form-meaning correspondence.

Experiment 2: Repetition of gesture form and the underlying representations

In this experiment we investigate whether a perceived gesture form can influence the construction of meaning (whether it be any semantic representation or a conceptual pact), which subsequently influences gesture production. To test this, we have used a route description task. By asking a participant and a confederate

to give each other directions repeatedly, a situation was created in which it is quite natural to repeat each other's gesture forms, without drawing much attention to this process. Because of the more interactive setting, both low-level and high-level processes (e.g. audience design) that may be involved in the repetition of gesture forms across interlocutors can come into play.

Giving directions allows for different conceptualizations of the task at hand. We presented participants with bird's view drawings of a city scene, which had a short route indicated on them (see Figure 4 for an example). These scenes were neither presented vertically nor horizontally, but at an angle. Therefore, the production task could be thought of as either describing a route on a vertically oriented map, or as describing a route through an actual (horizontal) city.

The confederate always was first to give a route description. Although her verbal descriptions were the same across conditions, her gestures differed. The movement of her gestures was either in accordance with the conceptualization of indicating the route on a vertically oriented map, which we call *Vertical Map* perspective, or with the conceptualization of a route through a city, which we call *Route* perspective. Thus, two different conceptualizations were suggested in gesture.

It is interesting in itself to see whether participants adapt to the confederate's perspective in gesture. Yet this alone would not tell us whether this is based on a direct repetition of gesture form, or on the convergence of representations of meaning. Therefore, apart from manipulating perspective, we manipulated hand

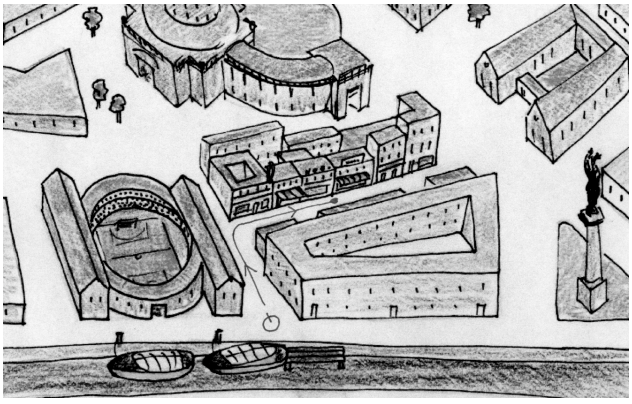

Figure 4: Part of a city scene used in the experiment,
note the route starting at the bottom-center.

shape independently. Our intuition was that when describing a route through a city, hand points, where all fingers are extended as an index, as well as finger points, where only one finger is extended as an index can be used, whereas when pointing on a map it is more common to point with one finger than with all fingers extended.

We tested this intuition in two ways: by analyzing pictures on the Internet and by asking people in the streets for directions. First, we did a Google search for pictures on the bigram "getting directions" on December 14th 2010 in the Netherlands. We selected those photographs in which a person was pointing, seemingly to communicate to another person, and whose pointing hand was clearly visible. We scored photographs on the first five pages of search results for whether the person in the picture was pointing at a map or not and whether they were pointing with their hand or with their finger. Duplicate search results were counted only once. We found 6 finger points at a map, 0 hand points at a map, 10 finger points that were not on a map, and 8 hand points not on a map, in line with our hypothesis. Next, we searched for more gestures pointing at a map with the term "pointing at a map". We found 50 finger points at a map, 1 hand point at a map, and 7 points at a map with another index, such as a pen or a stick. These data support our hypothesis that finger pointing is the most common way of pointing at a map.

Additionally, we asked 20 Dutch-speaking adults in the city center of Tilburg for directions, while holding a paper map in hand. Out of the gestures that were pointing at the map, 29 were produced with one finger as an index, while 1 was produced with the tip of a key. Out of the gestures that were pointing into the streets, 18 were produced with one finger as an index and 20 were produced with all fingers extended. We excluded the 'key-gesture' from our analysis and found that this distribution is unlikely to be accidental, Yates $\chi^2(1) = 19.32$, $p < .0001$. Thus, our hypothesis that people are less likely to point with all fingers extended when pointing at a map was supported by the data.

We use this difference in common hand shapes between the domain of giving directions using a map and the domain of giving directions in the streets to address our research question. If it is the case that gesture form is perceived and reproduced directly, without the conceptual level being involved, participants may adapt to any of the gesture features produced by the confederate. That is, they may adapt to the confederate's hand shape and to the confederate's perspective. One of these may be perceived more easily than the other, so there

could be a difference in the extent to which each of the features is adapted to, but what we would not expect based on this view, is for the confederate's perspective to influence a participant's hand shape or for the confederate's hand shape to influence the participant's perspective.

On the other hand, if meaning does form an intermediate stage between the perception and production of a gesture form, we do expect such cross-effects to occur. For example, our pre-test showed that it is more common to point at a map using a single finger, than it is to point at a map using four fingers at once. Therefore, if the confederate's vertical movements would lead participants to think of this task as describing a route on a map, their gestures may be more frequently produced with one finger as an index as opposed to four. This would mean there is an effect of the perspective of the confederate's gestures on the hand shape of participants' gestures. This effect may also be found in the reverse direction: the use of all fingers as an index may lead participants to more readily think of the route as through a city than on a map, causing them to gesture in the Route perspective rather than the Vertical Map perspective.

Method

Participants Forty-eight participants took part in this experiment, out of which we excluded 6 from our analysis because they did not produce any of the gestures we were interested in (path gestures) and 2 because they indicated some suspicion about the experiment (see *Procedure* and *Statistical analysis* below). We used the data of 40 (33 female) participants in our analysis.

Procedure The participant and the confederate came to the lab and were introduced by the experimenter. They each received a written instruction and were seated across from each other. The instruction explained a communication task, and stated that the couple with most correct responses could win a book voucher (in reality there was a random draw). To their side (right to the participant) was a table, on which sat a flip chart for each interlocutor. In between these flip charts was a screen, such as to keep information private. The screen did not keep the interlocutors from seeing each other. Both behind the confederate and behind the participant was a camera capturing the other interlocutor. After reading the instruction, both 'participants' were allowed to ask questions. The confederate always asked one question, after which the

experimenter quickly went over the task again. Then the experimenter turned on the cameras and left the room.

The confederate started by studying a little map and memorizing the route on it. Each route had one turn, see Figure 4 for an example. She then turned the page of her flip chart (rendering a blank page) and described the route to the participant, for example: "Je begint bij de rondvaartboot, dan ga je langs het voetbalstadion en dan rechts een winkelstraat in tot ongeveer halverwege." ("You start at the tour boat, then you go along the soccer stadium and then into a shopping street on the right until about halfway.") The confederate's speech followed a script and was the same in each condition. The terms used to describe the directions were consistent with a horizontal perspective such as 'rechtdoor' (straight ahead) and 'steekt over' (cross), or were neutral for perspective 'tot ongeveer halverwege' (until about halfway). Gestures were timed naturally with speech and gazed at by the confederate. The confederate gestured with her right hand. The first direction of a route was always straight, which was depicted with either a forward or an upward movement. These movements were of comparable size. The gesture for the second direction (to the side) was placed relative to the first gesture; it started where the first gesture had ended.

After the confederate's description, the participant turned a page and was to choose which route had just been described, selecting from four alternatives by pronouncing the corresponding letter, see Figure 5. No feedback was provided. Then it was the participant's turn to study a route. This route was always on the same scene that the confederate's route had been on. After turning the page (rendering a blank page) the participant described the route to the confederate, who then turned a page and selected one of the four alternatives. This ended one cycle of the experiment. In total each participant perceived and produced five route descriptions, which took between six and eleven minutes. The confederate's descriptions took about twelve seconds each. On average, participants took about equally long for their descriptions, ranging from seven to nineteen seconds. (Most time of the experiment was filled with studying the maps and selecting answers.)

Afterward, both the confederate and the participant filled out a questionnaire, which included questions on the presumed purpose of the experiment and whether the participant noticed anything peculiar. Participants were also asked if they had recognized the city in the pictures, which none of them had (the drawings were loosely based on St. Petersburg). When the participant was done

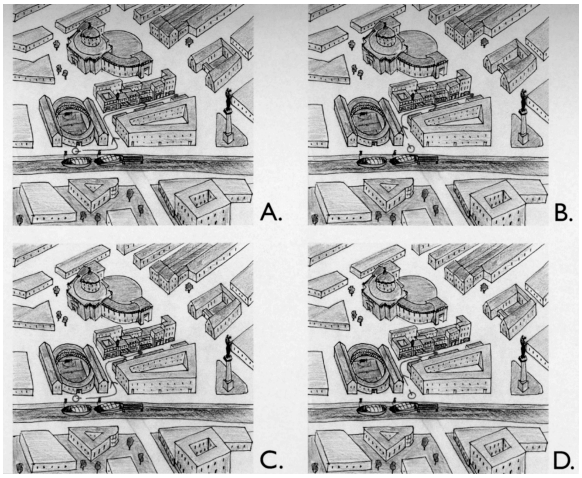

Figure 5: Example of the routes to choose from. Each map has a slightly different route depicted on it. Participants selected a route by calling out the corresponding letter.

filling out the forms, the confederate revealed her role and asked the participant's consent for the use of their data. Participants were asked if they had suspected any deception. The data of two participants was excluded from our analysis, because they indicated having been suspicious about either the goal of the experiment or the role of the confederate.

Design We have used a 2 x 2 between subjects design. The independent variables were the perspective (Route or Vertical Map) and hand shape (one or four fingers extended) of the confederate's path gestures. In the Route perspective, gestures were performed in the horizontal plane, with the index in the direction of the hand movement, as though following a virtual route (Figure 6a, 6b). In the Vertical Map perspective, gestures were performed in the vertical plane and the index was always pointing forward, as though pointing on a virtual map (Figure 6c, 6d).

Coding We coded all *path gestures* that participants produced, that is, all gestures in which one or more fingers were extended as an index, there was hand movement along some virtual path, and the co-occurring speech mentioned a direction to take. Within the stroke phase of each path gesture, we coded hand shape and perspective. The labels for hand shape were *Finger*, when one finger

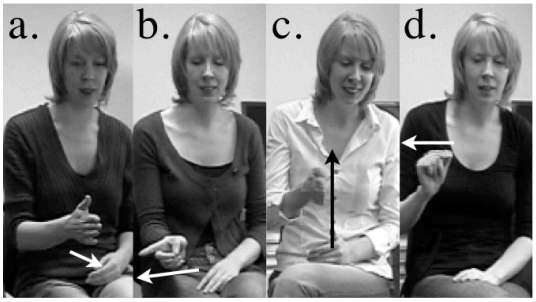

Figure 6: The confederate's path gestures. a: Hand/ Route; b: Finger/ Route; c: Hand/ Vertical Map; d: Finger/ Vertical Map.

was extended as an index, and *Hand*, if more than one finger was extended. The label for perspective was based on the following features of participants' gestures: location in the gesture space, hand orientation, and movement (direction and size). The label that could explain most features was assigned to each gesture. It turned out that in addition to the two perspectives that the confederate had used, participants occasionally used an alternative one, as though pointing on a horizontal map. Therefore, we chose from three labels: *Vertical Map*, *Route*, and *Horizontal Map.* A gesture in the Vertical Map perspective typically has vertical movement, with relative sizes mapping onto distances on the map, fingers pointing forward and the location in the gesture space corresponding to the location on the map (Figure 7a, 7b). Route gestures on the other hand have horizontal movement in front of and to the side of the speaker, with the fingers pointing in the direction of the hand movement (Figure 7c). Horizontal Map gestures (Figure 7d) differ from Route gestures in their hand orientation (fingers pointing down), and their relative size and location. Figure 7 shows some examples of participants' path gestures and our coding.

Initially all gestures were coded by a single coder. Reliability was assessed by having a second coder code a random sample of 20% of the path gestures in each condition for hand shape and perspective, N = 58. The two coders agreed on the label for hand shape in 95% of the cases, Cohen's Kappa = .89, indicating almost perfect agreement (Landis & Koch, 1977). Given the observed marginal frequencies of the labels, the maximum value of Kappa was .96. The two coders agreed on the label for perspective in 79% of the cases, Cohen's Kappa = .66,

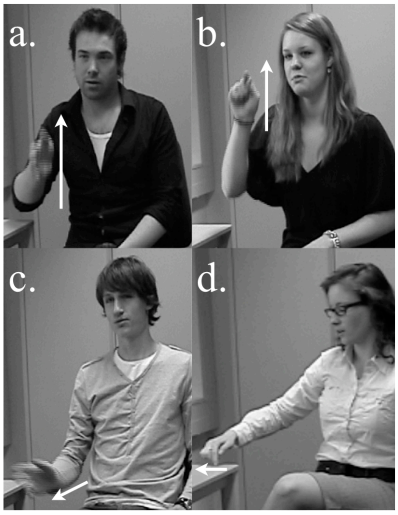

Figure 7: Examples of participants' path gestures and our coding. a: Hand/ Vertical Map; b: Finger/ Vertical Map; c: Hand/ Route; d: Finger/ Horizontal Map.

indicating substantial agreement (Landis & Koch, 1977). The maximum value of kappa was .86 in this case. In our analyses, we used the coding of the first coder.

Statistical analysis Analyses were done using ANOVA, with factors perspective (levels: Vertical Map, Route) and hand shape (levels: Hand, Finger) of the confederate's gestures. There were 40 participants, 10 in each cell. As a measure of participants' perspective, we report the number of path gestures that a participant produced in the Vertical Map perspective divided by all path gestures produced by that participant. As a measure participants' hand shape, we report the number of gestures that a participants produced with one finger extended, divided by the total number of path gestures produced by that participant. The significance threshold was .05 and we report partial eta squared as a measure of effect size.

Results

Participants' perspective Analyses of the perspective of participants' gestures, shown in Figure 8, revealed a main effect of the confederate's perspective, such that participants produced a larger proportion of Vertical Map gestures when the

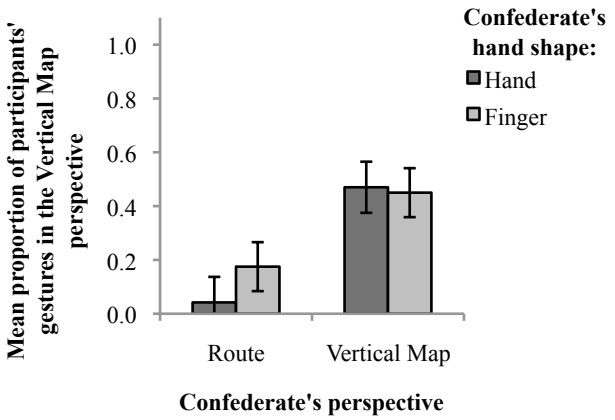

Figure 8: Mean proportion of gestures that participants produced in the Vertical Map perspective for each perspective and hand shape used by the confederate. Bars represent standard errors.

confederate gestured in the Vertical Map perspective (M = .46, SD = .35) than when the confederate gestured in the Route perspective (M = .11, SD = .20) (M_D = .35, 95% CI = .17, .54), $F(1, 36) = 14.88$ $p < .001$, η^2_p = .29. The confederate's hand shape did not exert a main effect on the perspective of participants' gestures, as they produced about equal proportions of Vertical Map gestures when the confederate gestured with all fingers extended (M = .26, SD = .34) and when she gestured with one finger extended (M = .31, SD = .33) (M_D = -.06, 95% CI = -.24, .13), $F(1, 36) = .38$, p = .54. These two factors did not interact, $F(1, 36) = .71$, $p = .41$.

Participants produced Horizontal Map gestures in about equal proportions across conditions. Therefore, the results for participants' gestures in the Route perspective mirror the results reported above.

Participants' hand shape Analyses of the hand shape of participants' gestures, shown in Figure 9, revealed a main effect of the confederate's perspective, such that participants produced a larger proportion of gestures with one finger extended when the confederate gestured in the Vertical Map perspective (M =.48, SD = .43) than when the confederate gestured in the Route perspective (M = .22 , SD = .37), $F(1, 36) = 5.00$, $p < .05$, η^2_p = .12. The confederate's hand shape did not exert a main effect on participants' hand shape, as participants

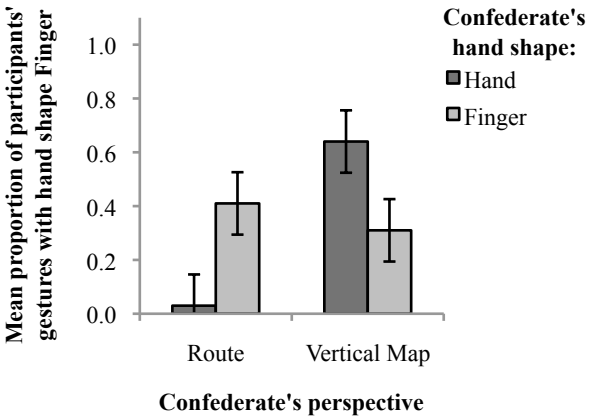

Figure 9: Mean proportion of gestures that participants produced with one finger extended (hand shape Finger) for each perspective and hand shape used by the confederate. Bars represent standard errors.

produced about equal proportions of gestures with one finger extended when the confederate gestured with all fingers extended (M = .34, SD = .39) and when the confederate gestured with one finger extended (M = .36, SD = .45) (M_D = -.02, 95% CI = -.26, .21), $F(1, 36) = .04$, $p = .85$. These two factors interacted, $F(1, 36) = 9.40$, $p < .01$, $\eta^2_p = .20$: When the confederate gestured in the Route perspective, participants adapted to her hand shape, as they produced a larger proportion of gestures with one finger extended when the confederate gestured with one finger extended (M = .41, SD =.45) than when she gestured with all fingers extended (M = .03 , SD = .07) (M_D = .38, 95% CI = .08, .68), $F(1, 18) = 6.94$, $p < .02$, $\eta^2_p = .28$. Yet when the confederate gestured in the Vertical Map perspective, participants did not adapt to her hand shape, as they produced a larger proportion of gestures with one finger extended when the confederate gestured with all fingers extended (M = .64, SD = .32) than when she gestured with one finger extended (M = .31, SD = .48), $F(1, 18) = 2.24$, $p = .16$.

Discussion

As predicted by both the theory that gesture form is copied directly and theories that representations of meaning are involved, participants adapted to the perspective used by the confederate. When the confederate gestured in the Vertical Map perspective, participants were more likely to do so as well and

similarly for the Route perspective. But was this merely a copying of form, or were they adapting to the conceptualization expressed in the confederate's gestures?

Importantly, we found some of the cross-effects we expected if perceiving certain gestures would cause participants to think of the task in a certain perspective, which in turn would influence their own gesture production. The perspective of the confederate's gestures influenced the hand shape of participants' gestures: participants more frequently pointed with one finger if the confederate gestured as though on a vertical map. This can be explained as the confederate's vertical gestures leading participants to think of the route as on a map, which caused them to point with their finger. If this is so, we would expect this increase in the use of one finger being caused by gestures that participants produced with the Vertical Map perspective, rather than in the Route perspective. This was indeed the case. Participants produced fewer gestures with one finger in the Route perspective when the confederate gestured in the Vertical Map perspective, than when she gestured in the Route perspective. It thus seems that participants were not merely repeating the individual features of the confederate's gestures, but rather the meaning that they expressed.

A theory that only takes into account the alignment of gesture forms can also explain that participants adapted to the confederate's perspective. Yet such a theory would not predict participants to gesture with one finger as an index more frequently when the confederate gestured in the Vertical Map perspective, particularly not when the confederate gestured with all fingers extended. However, it is possible to explain this effect in terms of biomechanics. For example, it may be the case that vertical hand movements lead people to extend their index finger, rather than all fingers, simply because it is easier, without them associating this with pointing at a map at any level of processing (our theory is neutral as to whether conceptual representations are more embodied or more symbolic in nature). This would be an explanation without conceptual mediation. However, we do not know of any data that support a theory that it is easier to extend one finger instead of all fingers when lifting an arm. An informal pilot study showed that when people were asked to copy the confederate's gestural movements, without any meaning being ascribed to them, people could do so effortlessly, both in terms of movement and hand shape. In our view, this makes an explanation in terms of biomechanics less likely. Also,

rather than explaining the data post hoc, our theory predicts the effect we found, making it more powerful.

Overall, perspective was adapted to more than hand shape. This may be because perspective was expressed in two features (movement and hand orientation), whereas hand shape is only one feature. Thus, the one non-matching feature may have been adapted to the two matching ones. It may also be because in this task, perspective carried a more critical meaning than did hand shape. A vertical gesture cannot possibly depict a route one could walk (at least not in the Netherlands), whereas the distinctions between the different hand shapes seem far subtler. In other words, in this task, the perspective of gestures may have given rise to (shared) conceptualizations more readily than the hand shape with which they were produced.

Although we did not find an overall effect of hand shape, participants did adapt to the confederate's hand shape in the conditions in which she gestured horizontally in the Route perspective, whereas there was no adaptation to the confederate's hand shape in the Vertical Map conditions. A possible explanation for this is again one in terms of meaning. The different hand shapes may be more readily interpreted in a meaningful way when gesturing as though along a horizontal route than when gesturing as though on a vertical map, better allowing participants to construct concepts that were consistent with the confederate's hand shape.

General discussion and conclusion

Our experiments have shown that adaptation in representational gestures resembles adaptation in verbal references in various ways. First, certain gesture forms were more likely to be used after they had been perceived. Participants who saw a speaker in a stimulus movie produce certain representational gestures were more likely to produce these gestures later on, while retelling the speaker's story. Second, gesture forms were only repeated across speakers if they had occurred in a meaningful context. That is, if the gesture form could be interpreted in light of the meaning expressed in the concurrent speech. Lastly, there were instances where gesture form was not copied in a low-level automated way, but rather similar forms were used to express similar meanings, and aspects of a form that did not match a meaning were not copied but rather adapted to the

meaning. These findings go well with theories and models in which gesture and speech both stem from a single concept, idea, or communicative intention (e.g. Cassell, et al., 1998; De Ruiter, 2000, 2007; Kendon, 2004; Kita & Özyürek, 2003; McNeill, 1992).

In experiment 1B, when a perceived gesture was incongruent with the content of the accompanying speech, the gesture was not repeated when the speech content was. In terms of the interactive alignment account, this may be because the incongruent gesture form that was perceived did not match the representation of meaning that was formed in the interpretation process, and thus no link was established between the representation of meaning and a representation of gesture form. Therefore, when this representation of meaning was subsequently activated by the production process, the gesture form was not. This explanation also fits the one exception we found, where an incongruent gesture was repeated, but the content of speech in the retelling differed markedly from the original story. In this case, the interpretation formed during perception seems to have incorporated the incongruent gesture. Therefore, the representation of meaning activated during production could activate the gesture form that had been perceived, but not the by then incongruent lexical forms that were perceived. This one case is very similar to the results found by Cassell et al. (1998), who accounted for their results similarly.

Interestingly, participants were less likely to produce any gesture at all while expressing content that had been presented to them with an incongruent gesture. It may be that the perception of an incongruent gesture disturbed the activation of representations of the spatial and motor aspects of the event described, thereby making it less likely that a gesture was produced while retelling this event (also see Kelly, Özyürek, & Maris, 2010). In terms of the interface model (Kita & Özyürek, 2003), this can be explained as the gesture generator not being able to retrieve relevant data from working memory, and thus not being able to generate a gesture form.

Consistently, the gesture as simulated action framework (Hostetter & Alibali, 2010) would also predict that speakers are less likely to gesture when describing an event that does not involve the perception or performance of a particular action. This framework explains the production of representational gestures as simulating action as part of thinking for speaking. Therefore, it might predict that not having perceived a gesture congruent with the upcoming speech would cause participants to be less likely to produce a representational gesture. Alternatively,

it may also be that participants omitted the gesture for social reasons. It is known that adaptation has positive social consequences (e.g. Van Baaren, Holland, Steenaert, & Van Knippenberg, 2003), thus it may also be that producing a completely different gesture sends a negative social message, which participants may have wanted to avoid.

In addition to the first two experiments, the results of our route directions experiment also suggest that concepts underlie the repetition of gesture forms across interlocutors. Participants readily adapted to those features of a confederate's gestures that could be interpreted meaningfully, such as whether the gestures were produced horizontally, as though walking through a city, or vertically, as though pointing on a vertically oriented map. However, features that were inconsistent with these conceptualizations, notably the use of four fingers while gesturing as though pointing on a map, were not adapted to. Instead, we saw an effect of the confederate's perspective in gesture on the hand shape used by participants: if the confederate gestured as though on a map, participants more frequently used one finger as an index, which is consistent with the conceptualization of pointing out the route on a map.

Both the theory that interlocutors form conceptual pacts (Brennan & Clark, 1996) and the interactive alignment account (Pickering & Garrod, 2004) can account for these findings. In the account by Brennan and Clark, speakers adapt to each other's gestures with the aim of arriving at a shared conceptual understanding (also see Holler & Wilkin, 2011). Although plausible, such a functional account is not required to explain our data. The interactive alignment account can do so as well. Yet adaptation at one linguistic level alone cannot explain why specifically those aspects of a perceived gesture that could not readily be interpreted meaningfully were not reproduced. Instead, these features tended to be produced such that they fitted an interpretation consistent with most aspects of the perceived gesture. This can be explained using the links between different levels within a speaker, that is, between representations of meaning and representations of form. Only those features of a form that can be linked to a representation of meaning during interpretation are activated through this same representation of meaning once it is activated for production.

Regarding proposed models of speech and gesture production, such as models following the postcard architecture (De Ruiter, 2007) and the interface architecture (Kita & Özyürek, 2003), our results seem to emphasize that it is important for such models to have gesture and speech production and

interpretation share representations of meaning at some level. Both types of models allow for this at the conceptual level. Currently, these architectures do not provide an explicit account of adaptation in either gesture or speech, because they do not specify how production and perception are linked, or at what level representations are shared between these processes. In future work we intend to further study adaptation in gesture, adaptation in speech, and their possible influence on each other to shed more light on this issue. A model that can account for our current findings will need to allow for the conceptual level to influence what features of a perceived gesture form will be more likely candidates for gesture production.

As explained above, our results fit well with the theory that when communication partners interact, the concepts of both interlocutors converge and certain forms are used to refer to these shared concepts (Brennan & Clark, 1996; Garrod & Anderson, 1987). However, our results do not provide evidence that this is a deliberate process, or that it is part of audience design. Especially in our first two studies, audience design does not seem the most likely explanation, since even though the addressee was not the person who produced the original gestures, some of the original gesture forms were repeated when narrating to this new addressee. The convergence of representations of meaning may also happen automatically (Pickering & Garrod, 2004), without conscious effort or intent (but see Brennan & Hanna, 2009). This issue needs to be addressed in future research. Additionally, we have not yet studied gesture in a setting in which conceptual pacts can be arrived at incrementally. Rather, we have made use of a stimulus movie or a confederate whose gesturing followed a script, such that adaptation could only happen one way. Despite these limitations, our results suggest that the perception of meaningful forms in gesture can contribute to the convergence of concepts across interlocutors, which in turn informs gesture production.

Our results do not imply that features of gesture forms are never repeated without representations of meaning being involved. So far, we have only studied certain representational gestures. Our results may not generalize to other types of gestures, especially non-representational gestures, whose repetition across interlocutors may be more similar to that of other behaviors not carrying propositional meaning. Yet we have shown that certain representational gestures are only repeated if they make sense in the linguistic context and that one aspect of a perceived gesture form (perspective) can influence another aspect (hand

shape) of a gesture form produced. These results suggest that it is sometimes fruitful to include representations of meaning in an explanation of adaptation in non-verbal language use, especially when these behaviors carry propositional meaning. Rather than perceiving a form leading to the production of that form directly, we have shown that for representational gestures, meaning can play a mediating role. That is, representations of meaning are also converging across interlocutors rather than just representations of form, and this convergence of meaningful representations may be driving adaptation in gesture.

Acknowledgements

We thank all participants to these studies. We thank the anonymous reviewers and the editor of the Journal of Memory and language for their insightful comments to earlier versions of this work. We gratefully thank Susan Brennan and Sotaro Kita for valuable discussions on earlier versions of this work. We thank Nathalie Bastiaansen for drawing the stimuli of experiment 2, Els Jansen and Anouk van den Berge for collecting and coding the data of experiment 1, Vera Nijveld for doing the reliability coding of experiment 2, and our co-workers at Tilburg University for assisting in the data collection.

References

Bavelas, J., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse Processes, 15*, 469-489.

Bergman, K., & Kopp, S. (2009). Increasing expressiveness for virtual agents - autonomous generation of speech and gesture. In K. Decker, J. Sichman, C. Sierra & C. Castelfranchi (Eds.), *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009)* (pp. 361-368). International Foundation for Autonomous Agents and Multiagent Systems.

Bock, J. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*, 355-387.

Branigan, H. P., Pickering, M. J., Pearson, J., & McLean, J. F. (2010). Linguistic alignment between humans and computers. *Journal of Pragmatics, 42*, 2355-2368.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology-Learning Memory and Cognition, 22*(6), 1482-1493.

Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialogue. *Topics in Cognitive Science (Special Issue on Joint Action), 1*, 274-291.

Cassell, J., McNeill, D., & McCullough, K.-E. (1998). Speech-gesture mismatches: Evidence for one underlying representation of linguistic & nonlinguistic information. *Pragmatics & Cognition, 6*(2), 1-33.

Chartrand, T. L., & Bargh, J. A. (1999). The Chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology, 76*(6), 893-910.

Chui, K. (2005). Temporal patterning of speech and iconic gestures in conversational discourse. *Journal of Pragmatics, 37*(6), 871-887.

Cleland, A. A., & Pickering, M. J. (2003). The use of lexical and syntactic information in language production: Evindence from the priming of noun-phrase structure. *Journal of Memory and Language, 49*, 214-230.

De Fornel, M. (1992). The return gesture. In P. Auer & A. di Luzio (Eds.), *The contextualization of language*. Amsterdam: John Benjamins.

De Ruiter, J. P. (1998). Gesture and Speech Production. Unpublished Doctoral Dissertation. University of Nijmegen.

De Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and Gesture* (pp. 284-311). Cambridge: Cambridge University Press.

De Ruiter, J. P. (2007). Postcards from the mind: the relationship between speech, imagistic gesture, and thought. *Gesture, 7*(1), 21-38.

Effron, D. (1941). *Gesture and environment*. Morningside Heights, NY: King's Crown Presss.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica, 1*, 49-98.

Enfield, N. J., Kita, S., & De Ruiter, J. P. (2007). Primary and secondary pragmatic functions of pointing gestures. *Journal of Pragmatics, 39*, 1722-1741.

Feyereisen, P., Van de Wiele, M., & Dubois, F. (1988). The meaning of gestures: What can be understood without speech? *Cahiers de Psychologie Cognitive, 8*, 3-25.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition, 27*, 181-218.

Holler, J., & Wilkin, K. (2011). Co-speech gesture mimicry in the process of collaborative referring during face-to-face dialogue. *Journal of Nonverbal Behavior, 35*, 133-153.

Hostetter, A. B., & Alibali, M. W. (2010). Language, gesture, action! A test of the gesture as simulated action framework. *Journal of Memory and Language, 63*, 245-257.

Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science, 21*(2), 260-267.

Kendon, A. (1988). How gestures can become like words. In F. Potyatos (Ed.), *Crosscultural perspectives in nonverbal communication* (pp. 131-141). Toronto, Canada: Hogrefe.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Kimbara, I. (2006). On gestural mimicry. *Gesture, 6*(1), 39-61.

Kimbara, I. (2008). Gesture form convergence in joint description. *Journal of Nonverbal Behavior, 32*(2), 123-131.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface

representation of spatial thinking and speaking. *Journal of Memory and Language, 47*, 16-32.

Kita, S., Özyürek, A., Allen, S., Brown, A., Furman, R., & Ishizuka, T. (2007). Relations between syntactic encoding and co-speech gestures: Implications for a model of speech and gesture production. *Language and Cognitive Processes, 22*(8), 1212-1236.

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language, 57*(3), 396-414.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Levelt, W. J. M. (1989). *Speaking*. Cambridge, MA: MIT Press.

McNeill, D. (1992). *Hand and Mind: What gestures reveal about thought*. Chicago and London: The University of Chicago Press.

McNeill, D. (2005). *Gesture and Thought*. Chicago and London: University of Chicago Press.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2009). The communicative import of gestures: Evidence from a comparative analysis of human-human and human-machine interactions. *Gesture, 9*(1), 97-126.

Müller, C. (1998). *Redebegleitende Gesten. Kulturgeschichte - Theorie - Sprachvergleich.* Berlin: Berlin Verlag.

Parrill, F., & Kimbara, I. (2006). Seeing and hearing double: The influence of mimicry in speech and gesture on observers. *Journal of Nonverbal Behavior, 30*(4), 157-166.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language, 39*, 633-651.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169-225.

Tabensky, A. (2001). Gesture and speech rephrasings in conversation. *Gesture, 1*(2), 213-235.

Van Baaren, R. B., Holland, R. W., Steenaert, B., & Van Knippenberg, A. (2003). Mimicry for money: Behavioral consequences of imitation. *Journal of Experimental Social Psychology, 39*(4), 393-398.

Chapter 5

Gesturing by aphasic speakers: How does it compare?

Abstract
We compared gesturing by aphasic speakers to that of healthy controls, who either were or were not allowed to speak, to see whether gestures' intelligibility increases or decreases in aphasia, and to assess whether the techniques employed by aphasic speakers for depicting in gesture resemble those of healthy controls. We used video clips of 25 aphasic speakers and 17 healthy controls performing two communication tasks on a clinical test (Scenario Test). We conducted a perception study (15 raters per cell) to assess the intelligibility of their gestures and we studied their gestural representation techniques by means of coding and analysis. We found that gestures by aphasic speakers were less informative than those of healthy controls, and that gestures by people with severe aphasia were less informative than those by people with moderate aphasia. Aphasic speakers also tended to use fewer gestural representation techniques (mostly relying on outlining and deictic gestures) than did healthy controls who were asked to use gesture instead of speech. Our results suggest that in aphasia, gesture tends to degrade with speech. This implies that the processes underlying speech and co-speech gesture production may be tightly linked or shared.

This chapter is based on:
Mol, L., Krahmer, E., Van de Sandt-Koenderman, M. (submitted). Gesturing by aphasic speakers, how does it compare? *Journal of Speech Language and Hearing Research.*

Introduction

Gesture and speech production

When speaking, people oftentimes produce hand gestures that are closely linked to their speech temporally (Chui, 2005), structurally (Kita & Özyürek, 2003), and semantically (e.g. McNeill, 2005). Some of these *co-speech gestures* depict things in an iconic or deictic way. For example, when asking a sales clerk for a sweater, gestures may indicate that we prefer a V-neck (e.g. by outlining its shape), a large front pocket (e.g. by putting our hands into an imaginary sweater), or one just like the one we see on the mannequin (e.g. by pointing at it). Both the production of speech and the concurrent production of hand gestures seem to be part of a speaker's communicative effort (Kendon, 2004).

Although different functions of co-speech gesture have also been recognized, such as facilitating speech production (Krauss, 1998; Krauss, Chen, & Gottesman, 2000) and supporting cognition (Melinger & Kita, 2007), much empirical evidence has been gathered for the idea that co-speech gestures are communicative and are intended as such (e.g. Alibali, Heath, & Myers, 2001; Beattie & Shovelton, 1999). This may mean that people with impaired verbal communication skills, like aphasic speakers, can use gesture to compensate for their speech impairment. On the other hand, since speech and co-speech gesture production seem to be closely related, it may also be the case that co-speech gesturing breaks down with speech in aphasia. In this chapter, we address the question of whether speech and gesture are two sides of the same coin (McNeill, 2005), or whether they can compensate for one another (e.g. Butterworth & Hadar, 1989). We do so by assessing whether co-speech gesturing tends to be impaired in aphasic speakers.

Not all gestures are coordinated with speech. There are also gestures that are produced and understood without concurrent speech. One category of such gestures is formed by *pantomimes* (Kendon, 1988). Pantomimes are gestures whose form does not carry a conventionalized meaning. In this sense, they are similar to co-speech gestures, rather than to *signs* in languages of the deaf and *emblems* (e.g. the thumbs-up gesture in English), which do carry a conventionalized meaning. Yet similar to signs and emblems, pantomimes replace speech rather than accompanying it. Since pantomimes are not accompanied by speech, the process underlying pantomime production may not be linked as tightly to the process of speech production as may be the case for

co-speech gesture production. There is some evidence that the production of pantomimes relies on different resources than the production of co-speech gestures (e.g. Bartolo, Cubelli, Della Sala, & Drie, 2003; Goldin-Meadow, So, Özyürek, & Mylander, 2008; Rose & Douglas, 2003). For example, Goldin-Meadow et. al found that the predominant order in which agent, patient and action are mentioned in a speaker's native language while describing a motion event influences the order of expression in co-speech gestures, but not in pantomimes.

Because of the synchrony in time and meaning between speech and co-speech gestures, theories have been proposed on how speech and co-speech gesture production are interrelated. McNeill (2005) argues that speech and gesture co-express idea units, which develop themselves into utterances. That is, in the process of thinking-for-speaking, ideas are developed from their origin, which McNeill refers to as the *growth point*, into an utterance that consists of both speech and gesture. In this view, speech and gesture are two sides of the same coin. In support of this idea, So, Kita, and Goldin-Meadow (2009), for example, found that if information was lacking in speech, it tended to be missing in gesture as well.

Melinger and Levelt (2004) on the other hand, found that speakers sometimes divide the content of their message across gesture and speech. They found that if critical spatial information was expressed in gesture, it was more likely to be omitted in speech. This goes well with the idea that gesture and speech production are complementary and can compensate one another, which also underlies the Tradeoff Hypothesis. This hypothesis states that "when speaking gets harder, speakers will rely relatively more on gestures", and vice versa (De Ruiter, Bangerter, & Dings, In Press). However, De Ruiter et al. found only little evidence that people gesture more when speech is harder. Rather, they found that gesture and speech tended to express similar types of information, consistent with the idea that gesture and speech are two sides of a coin.

Gesture production and aphasia

In light of the question of whether gesture and speech can compensate for one another, it is interesting to study what happens to gesture when speech breaks down, such as in aphasia. Aphasia is an acquired language disorder caused by brain damage, affecting all language modalities. The severity of the language

disorder may vary considerably. In our current study we focus on aphasic people who have severe to moderate problems expressing themselves verbally.

Numerous studies have shown that aphasic people still gesture spontaneously and frequently (Rose, 2006). People with fluent aphasia may even produce more gestures that convey information than non-aphasic speakers (Carlomagno, Pandolfi, Marini, Di Iasi, & Cristilli, 2005). Case studies and clinical experience confirm that some aphasic speakers use gesture effectively to communicate (e.g. Goodwin, 2002). This suggests that people with aphasia may be able to partly compensate for their speech impairment with gesture. Yet does this mean their gesturing is unimpaired?

When producing a content-bearing gesture, there are different ways in which we can depict (Cienki & Müller, 2008). For example, if we want to depict a sweater, we can outline its shape, or we can pretend to put it on. And if we are talking about a car, we can move our hands as though steering it, or we can let our hand represent the car, depicting its path with our hand movement. When produced along with speech, each of these gestures would fall into the category of *iconic gestures*, which are gestures that mostly depict entities or movements (McNeill, 2005). Being able to produce a meaningful iconic gesture does not mean that all these different representation techniques are intact. Therefore, to know whether gesture is impaired in aphasia, we need to study both its meaning and its form, and we need to compare aphasic speakers to non-aphasic speakers.

Carlomagno and Cristili (2006) compared five speakers with fluent aphasia, five speakers with non-fluent aphasia and ten control participants for how well they could convey two pieces of news verbally and how they gestured. They found no evidence for a difference between the three groups in what features of objects and actions were depicted in iconic gestures. They did find that speakers with fluent aphasia produced more iconic gestures than the other groups, while the speakers with non-fluent aphasia produced more *deictic* and *metalinguistic* gestures than the other groups. Deictic gestures are for example pointing gestures. Metalinguistic gestures are gestures that do not refer to the same content as the concurrent speech, but rather show the structure of the spoken message or comment on it, for example expressing uncertainty. Carlomagno and Cristili conclude that for people with fluent aphasia, there can be a mismatch between their speech content, which results from their impaired processing of verbal semantics, and their gesture forms, whose production may be unimpaired. This was suggested earlier by Butterworth and Hadar (1989), and it is in line

with the view that gesture and speech can compensate for one another and thus are relatively independent.

Studying the relation between speech impairment and gesture, Cocks, Dipper, Middleton and Morgan (2011) drew a detailed comparison between gestures produced by a speaker (LT) with conduction aphasia and those of non-aphasic speakers. They found that LT's gestures during tip-of-the-tongue states (co-ToT gestures) differed from those accompanied with fluent speech by herself and the control speakers. For example, most of the co-ToT gestures outlined shapes, whereas this was rarely the case for co-speech gestures. Cocks et al. explained this result by the fact that the co-ToT gestures mostly seemed to refer to objects and animals, which may be easily depicted that way, whereas the co-speech gestures depicted events. Although this is a plausible explanation, there is an alternative account as well. McNeil and Duncan (2010), explain aphasia as a problem in translating intact conceptual knowledge into an utterance (consisting of speech and gesture). Therefore, when aphasic speakers have trouble verbalizing their idea, it may be foremost the conceptual knowledge that is hard to access, or to translate into an utterance. Producing shape-outlining gestures requires relatively little use of conceptual knowledge, as it is grounded in the perceptual features of the referent. Therefore, one does not need to know how to use the referent, or how the referent typically behaves in order to perform the gesture. Thus, producing shape-outlining gestures may largely surpass the processes of retrieving conceptual knowledge and of translating conceptual knowledge into an utterance. This may also explain in part why especially shape-outlining gestures were used during word finding problems.

Cocks et al. also found that the differences in LT's gesturing paralleled the differences in her speech, suggesting that although LT could still use gesture effectively, her gesture production was impaired, much like her speech production. They call for a study in which iconic co-speech and co-ToT gestures of a larger number of aphasic and non-aphasic speakers are compared in various types of discourse.

Present study

In our present study we compare the iconic and deictic gestures of a larger number of aphasic speakers to those of non-aphasic speakers. However, because of the scale of our study, our approach differs from the approach by Cocks et al. (2011). We look at co-speech gestures, that is, any gestures accompanied by

speech, without discriminating between co-speech and co-ToT gestures. Doing so, we address our research question accurately, because all gestures that are accompanied by speech sounds are likely to stem from the process of co-speech gesture production, this as opposed to gestures that are not accompanied by speech, some of which may result from a separate process for producing pantomimes. Therefore, we look at co-silence gestures separately.

To assess whether or not co-speech gesture production tends to be impaired in aphasic speakers, we compare both the meaning and the form of gestures produced by speakers with severe aphasia, speakers with moderate aphasia, and healthy control participants. We also asked the healthy control participants to communicate without using speech, using gesture instead. This gives us insight into how people with an unimpaired gesture production system would compensate for speech with gesture, which informs us on whether aphasic speakers can use gesture to compensate for their speech impairment as freely as people without aphasia.

First, we look at the intelligibility of gestures. If aphasic speakers compensate for speech with gesture, we expect their gestures will be more informative than those of non-aphasic speakers, who can rely on speech more. Also, the more speech is impaired, the more informative gesture will be. Alternatively, if speech and gesture are two sides of a coin, and therefore also break down together, the opposite is expected. That is, gestures of aphasic speakers will be less informative than those of healthy speakers, and the more impaired speech is, the more impaired gesture will be. We test this by means of three perception experiments, in which we separately assess the informativeness of the verbal and nonverbal communication of people with moderate and more severe aphasia and healthy control participants. For this purpose, the speakers perform two communication tasks, which differ in how difficult it is to express the information that needs to be conveyed verbally.

Second, we present a detailed analysis of the form of iconic and deictic gestures produced by aphasic speakers and control participants, zooming in on their representation techniques. If their gesturing is unimpaired, the techniques used by aphasic speakers may resemble the techniques used by non-aphasic speakers. If aphasic speakers compensate for speech with gesture, the techniques they employ may also be more similar to those of non-aphasics who are asked to communicate without speech. On the other hand, if their gesturing is impaired, this may affect some techniques more than others, and therefore aphasic speakers

may prefer different techniques than non-aphasic speakers or gesturers, and there may be differences in the techniques used by people with moderate and more severe aphasia. Specifically, the more severe aphasia, the more speakers may rely on techniques that do not require much use of conceptual knowledge, such as outlining shapes.

Perception experiments

Method

Material We used video clips of 25 native Dutch stroke patients with aphasia (16 male). Types of aphasia included: Global (7), Broca (3), Wernicke (3), Anomic (1), Conduction (1), and non-classifiable (6). For four patients the type of aphasia was not known. The mean age was 56.92 years, *SD* = 10.86, Range 37 – 71. The mean time post-onset was 25.56 months, *SD* = 40.16, Range 1 – 152. Details for each speaker can be found in Table 1. All patients gave their informed consent for the use of their data for research purposes.

The patients were performing an experimental version of the Scenario Test (Van der Meulen, Van de Sandt-Koenderman, Duivenvoorden, & Ribbers, 2009). This test measures a person's ability to functionally communicate. To compensate for their impaired verbal expressiveness, patients may use any alternative and augmentative means of communication, including gesture, to get their message across. We used data from two subtasks. In the sweater task, the patient is explained a scenario in which they are in a store and want to buy a sweater. The clinician talks about a sales clerk approaching and asking: "How may I help you?". The patient is then to communicate as though addressing the sales clerk, for example by saying: "I would like to buy a sweater". In the accident task, the information to be conveyed is more complex. The clinician explains a scenario in which the patient witnessed an accident, in which a car hit a biker. A police officer then approaches the patient asking: "What happened?". The patient is then to explain what took place, as though addressing the officer.

Apart from the videos of aphasic speakers, we also used video data of non-aphasic control participants, who were matched for age and performed the same test items with a trained tester. Their mean age was 54.06 years, *SD* = 11.09, Range 33 - 77. They were allowed to speak on one subtask (verbal control

participants) and were asked to communicate using gesture exclusively on the other (nonverbal control participants).

We cut out fragments of the videos of all people performing the two subtasks, starting right after the final question posed by the clinician, and stopping right after the speaker's first attempt at communicating the required information. Out of these fragments, we made three stimulus movies for our perception studies: one containing all fragments of aphasic speakers, one containing all fragments of the verbal control participants, and one with all fragments of the nonverbal control participants. For the aphasic speakers and the

Table 1: Data of aphasic speakers.

Participant number	Gender	Age (years)	Type of aphasia	Time past onset (months)	ANELT score	Group
1	F	43	Global	2	10	Severe
2	M	68	Global	86	10	Severe
3	M	71	Global	67	10	Severe
4	M	52	Global	4	10	Severe
5	M	67	Global	1	10	Severe
6	F	37	Wernicke	2	10	Severe
7	F	68	Unknown	3	10	Severe
8	M	56	Global	4	10	Severe
9	M	59	Wernicke	49	11	Severe
10	M	67	Global	3	13	Severe
11	M	64	Non classifiable	1	19	Severe
12	M	69	Non classifiable	73	19	Severe
13	F	44	Non classifiable	6	22	Moderate
14	F	51	Unknown	1	23	Moderate
15	F	40	Broca	97	28	Moderate
16	M	68	Non classifiable	4	29	Moderate
17	F	66	Wernicke	1	29	Moderate
18	F	50	Broca	47	31	Moderate
19	M	70	Non classifiable	1	31	Moderate
20	M	57	Unknown	3	34	Moderate
21	F	41	Unknown	23	34	Moderate
22	M	63	Non classifiable	1	38	Moderate
23	F	49	Conduction	1	39	Moderate
24	M	47	Broca	152	40	Moderate
25	M	56	Anomic	7	43	Moderate

verbal control participants, we created three versions of these stimulus movies: one with just the video image and no sound, one with sound and blank video, and one with both image and sound. The clips of the nonverbal control participants were video image only, that is, without sound.

Raters and task Raters were native Dutch students from Tilburg University, who had no expertise in gesture or aphasiology. They performed a forced choice task, in which they were asked to judge whether the person in each clip of the stimulus movie was communicating that they wanted to buy a sweater, or that they had witnessed a car accident. When applicable, their instructions stated that the speakers they were about to see had a speech disorder. We did three separate perception studies, with different groups of raters. In the first study, we used the stimulus movies of the aphasic speakers only. Raters saw the video clips without sound, heard the audio clips without video, or saw and heard the video clips with sound. The second perception study was similar, but with the stimulus movies of the verbal control participants instead. Finally, we also did a perception test with the stimulus movie of the nonverbal control participants.

Statistical analysis Based on the severity of their verbal communication disorder, the aphasic speakers were divided into two groups. For this purpose, we used their score on the Amsterdam Nijmegen Everyday Language Test (ANELT) (Blomert, Kean, Koster, & Schokker, 1994). The ANELT is similar to the Scenario Test, except that only verbal communication contributes to a speaker's score. In line with the ANELT and the Scenario Test, we chose 20 as the cut off point. The ANELT labels patients with scores lower than 20 as the most severe group, whereas the Scenario Test discriminates between speakers with an ANELT score above and equal to 20 and those below, providing separate norms for the latter. Thus, speakers with a score below 20 (out of 10 – 50) were labeled as speakers with severe aphasia and speakers with a score above 20 were labeled as speakers with moderate aphasia. There were 13 speakers in the moderate aphasia group and 12 in the severe aphasia group. Since we ran our perception experiments separately, we present three separate analyses of variance. For pairwise comparisons we used the LSD method, with a significance threshold of .05. Our dependent variable in each analysis is the ratio of correct answers to all answers, averaged over raters.

Results

Table 2 shows the means and standard deviations of the ratio of correct answers to all answers, for clips from each group of 'speakers', for either task, and for each modality in which they were shown to the raters. Performance at chance level would render a score of .5. We first present an analysis of the study with clips from the two groups of aphasic speakers. We performed an ANOVA with Group (levels: Severe aphasia, Moderate aphasia) and Task (levels: Sweater, Accident) as within factors and Modality of presentation (levels: Visual, Audio, Audiovisual) as a between factor. There were 15 raters in each cell, 45 in total.

All factors showed a main effect. There were more correct answers when judging speakers with moderate aphasia (M = .87, SE = .01) compared to speakers with severe aphasia (M = .68, SE = .01), $F(1, 42) = 240.63$, $p < .001$, η^2_p = .85. There were also more correct answers when judging clips from the accident task (M = .81, SE = .01) than from the sweater task (M = .75 SE = 01), $F(1, 42) = 8.83$, $p < .01$, η^2_p = .17. There were fewer correct answers with the visual presentation (M = .64, SE = .02), compared to the audio (M = .83, SE = .02) and audio-visual (M = .86, SE = .02) presentation, $F(2, 42) = 66.48$, $p < .001$, η^2_p = .76. The difference between the latter two showed a trend toward significance, p = .07.

Table 2: Means and standard deviations of the ratio of correct answers to all answers.

		Mean ratio correct per modality		
Group	Task	Visual	Audio	Audiovisual
Severe aphasia	Sweater	.53 (.16)	.68 (.12)	.67 (.14)
	Accident	.68 (.13)	.70 (.12)	.84 (.10)
Moderate aphasia	Sweater	.69 (.13)	.94 (.04)	.98 (.03)
	Accident	.66 (.13)	.98 (.03)	.96 (.03)
Verbal control	Sweater	.78 (.14)	1.0 (.00)	1.0 (.00)
	Accident	.74 (.11)	.99 (.03)	1.0 (.00)
Nonverbal control	Sweater	.95 (.06)	-	-
	Accident	.90 (.06)	-	-

There was a two-way interaction between Group and Modality, $F(2, 42) = 25.62$, $p < .001$, $\eta^2_p = .55$, and between Group and Task, $F(1, 42) = 15.84$, $p < .001$, $\eta^2_p = .27$. The three-way interaction between Group, Task and Modality was significant as well, $F(2, 42) = 5.70$, $p < .01$, $\eta^2_p = .21$. Posthoc analysis revealed that in the visual modality, speakers with moderate aphasia were judged correctly more often than speakers with severe aphasia on the sweater task, $F(1, 14) = 36.63$, $p < .001$, $\eta^2_p = .72$, but not on the accident task, $F < 1$, *n.s*. In the audio modality, speakers with moderate aphasia were judged correctly more often than speakers with severe aphasia on both tasks.

Our next analysis compares the judgment of clips from speakers with moderate aphasia to that of clips from the control participants when they were allowed to speak (verbal control group). We used an ANOVA with Task as a within factor and Group and Modality as between factors. For clips from aphasic speakers, there were 15 raters per cell, and for clips from non-aphasic speakers there were 16 raters per cell, summing up to 93 raters in total.

There was a main effect of Group, $F(1, 87) = 12.14$ $p < .001$, $\eta^2_p = .12$. There were more correct answers when judging clips from the verbal control participants ($M = .92$, $SE = .01$) compared to those of speakers with moderate aphasia ($M = .87$, $SE = .01$). There also was a main effect of Modality, $F(2, 87) = 162.30$, $p < .001$, $\eta^2_p = .79$. Fewer speakers were judged correctly in the visual modality ($M = .72$, $SE = .01$), compared to the audio ($M = .98$, $SE = .01$) and audiovisual modality ($M = .99$, $SE = .01$). The interaction between Group and Modality was not significant, $F = 1.75$, $p = .18$, $\eta^2_p = .04$. There was a two-way interaction between Modality and Task, $F(2, 87) = 3.91$, $p < .05$, $\eta^2_p = .08$. In the visual modality, speakers were judged correctly slightly more often on the sweater task, whereas in the audio modality they were judged correctly more often on the accident task. Posthoc analysis confirmed that in the visual modality, the difference between the two groups of speakers was significant on both tasks, with the verbal control participants being judged correctly more often than the speakers with moderate aphasia.

Lastly, we present an analysis comparing the judgment of visually presented clips of the control participants when they could speak and when they could not speak (nonverbal control group). There were 16 raters in each cell, 32 in total. Task was again the only within factor.

There was a main effect of Group, $F(1, 30) = 24.84$, $p < .001$, $\eta^2_p = .45$. There were more correct answers when judging clips of nonverbal control

participants (M = .93, SE = .02) compared to clips of verbal control participants (M = .77, SE = .02). We did not find a main effect of Task, $F < 1$, *n.s.*, but there was an interaction between Group and Task, $F(1, 30) = 8.85$, $p < .01$, $\eta^2_p = .23$. For verbal control participants, more speakers were judged correctly on the sweater task whereas for nonverbal control participants more speakers were judged correctly on the accident task.

Discussion

Overall, clips from speakers with severe aphasia were judged less accurately than those of speakers with moderate aphasia, which in turn were judged less accurately than clips from the verbal control participants. This is not surprising when it comes to the audio modality. Yet we see the same trend for the visual modality. The verbal control participants were judged more accurately than the aphasic speakers on both tasks. Especially on the sweater task, clips from speakers with moderate aphasia were judged more accurately than clips from speakers with severe aphasia. Therefore, it seems that the aphasic speakers were not able to compensate for their verbal impairment nonverbally. This indicates that nonverbal communication breaks down with verbal communication, rather than it taking on the role of verbal communication.

The almost perfect scores on the clips of nonverbal control participants show that, in principle, nonverbal communication can largely compensate for speech on this simple judgment task. Seeing a speaker of course provides more information than just gestures. We think however that gesture was the most important nonverbal cue in our clips. It therefore seems that people with aphasia cannot use gesture as freely as people without aphasia to compensate for speech. Yet was their gesturing informative at all? On the sweater task, raters were performing at chance level when only seeing the severe aphasic speakers, indicating that their gesturing was not informative. Yet on the accident task, the gestures of speakers with severe aphasia did seem to provide some information, even when added to the audio modality, as can be seen from the higher score on the audiovisual than on the audio only modality. This shows that information in gesture and speech was not fully redundant. For some severe aphasic speakers, seeing them too was apparently more informative than just hearing them. This indicates that gesture did take on some of the communicative burden, especially for the severe aphasic speakers on the accident task. Like their speech, their

gestures still contained some information, even though their ability to gesture seems impaired.

Our results show that the gestures produced by aphasic speakers were less informative than those produced by healthy control participants, both when the controls were speaking and when they were not. Next, we take a closer look at the form of the gestures produced by aphasic and non-aphasic speakers. Would impairment in gesture lead to the use of different gesture forms?

Gesture analysis

Method

Gesture coding We coded all hand movements that seemed relevant to the communication task in each of the clips used in our perception studies. We currently focus on iconic and deictic gestures (McNeill, 2005). Deictic gestures included gestures locating objects in the gesture space and pointing gestures. Based on work by Müller (2008), we further coded all iconic gestures into three categories, based on the representation technique used to depict. Gestures that outlined something in the gestures space, either by showing its contour (2D) or molding its shape (3D) were labeled as *outlining/molding*, for example, drawing the outline of a sweater in the air. Gestures that depicted the handling of a virtual object, such as holding the hands up as if using a steering wheel to depict a car, were labeled as *handling*. Gestures in which the hands represented an object, or in which the entire body depicted the body of another person were labeled as *object/enacting.* Examples are moving an upright hand forward and then flipping it horizontally, to depict that a biker fell. Although theoretically possible, we found it too opaque to code deictic gestures into these categories. Therefore, such gestures were only labeled as *deictic.*

To assess reliability, a second coder examined 58 randomly selected gestures (15%) and coded for which of the four representation techniques was used in each gesture. Agreement between the two coders was 93%, Cohen's Kappa = .90, indicating almost perfect agreement (Landis & Koch, 1977). Given the relative distribution of the labels, the maximum value for Kappa that could have been obtained was .94. In case of disagreement, we used the coding of the first coder, to ensure consistent coding throughout the entire data set.

It is important to note that the applied coding scheme does not start from the interpretation of a gesture, but rather from its form. In most cases, it is possible to determine the technique used entirely from the form of a gesture, without knowing its meaning. For example, one can usually see whether a hand is drawing, grasping or representing, even if what it is drawing, grasping or representing remains subjective. If two speakers are both using outlining, and both these speakers intend to gesture about the shape of a sweater, this does not mean that they are equally successful in conveying this meaning, but rather that they choose the same technique to depict the sweater. Therefore, in the following section, when we speak of similarity we are talking about similarity in the techniques used to depict, not in the appearance or the effectiveness of the gestures.

We also coded whether or not each gesture co-occurred with speech. Gestures that were accompanied by speech or speech-like sounds were labeled as co-speech gestures, whereas gestures that were performed during silence were labeled as co-silence gestures. For the aphasic speakers and verbal control participants, even though some gestures did not co-occur with speech, they are not necessarily pantomimes. Pantomimes are likely to originate from a different production system than co-speech gestures. In some cases of the aphasic speakers, it seemed that the gesture had been intended as a co-speech gesture, but then suddenly speech stopped due to a word finding problem while the gesture was still executed. Thus, we cannot be sure whether these co-silence gestures resulted from the co-speech gesture production process or from the pantomime production process. Therefore, we analyze them separately from the co-speech gestures.

Statistical analysis We are interested in the extent to which the different representation techniques are used by each group. Since there may be differences in the number of gestures produced by each participant and each group, we look at the proportions of gestures of each type, rather than at the number of gestures of each type. Thus, we calculate the number of gestures a participant produced of a certain type, divided by the total number of representational gestures produced by that participant. To put these proportions into perspective, we first report the mean number of gestures that each group produced with and without speech. Next, we report the mean proportions of each gesture type for co-speech gestures. The nonverbal control participants only produced co-silence gestures.

Still, it is interesting to compare aphasic speakers' co-speech gestures to gestures of healthy speakers who compensate for speech with gesture, to see to what extent aphasic speakers are using the representation techniques that healthy control participants would choose if they had to express themselves through gesture instead of speech. The pantomimes of the nonverbal controls are therefore included in our analyses of co-speech gestures. Since the number of co-silence gestures produced by the other groups was very small, we only report descriptive statistics for the types of co-silence gestures they produced.

To assess whether there were differences between the groups in the representation techniques used in their co-speech gesturing, we conducted 4x2 ANOVAs with Group (levels: Severe aphasia, Moderate aphasia, Verbal control, Nonverbal control) and Task (levels: Sweater, Accident) as fixed factors. However, the two different tasks, with different information to be expressed, may inherently call for different representation techniques. Therefore, we also performed separate analyses for each task, independent of whether the interaction between Group and Task was significant. Pairwise comparisons were done using the LSD method, with a significance threshold of .05.

Results

Before looking at the different gesture types produced, we first present data on the total number of gestures produced. Table 3 shows the mean number of co-speech gestures and co-silence gestures/ pantomimes produced by each group on either task, using any of the four representation techniques. Since the nonverbal control participants produced far more gestures than any other group, we report a comparison of the other three groups (Severe Aphasia, Moderate Aphasia, Verbal Control) for the number of co-speech and co-silence gestures produced. In total, these three groups produced 217 co-speech and 27 co-silence gestures. The nonverbal controls produced 133 gestures.

There was a main effect of Group on the total number of co-speech gestures produced, $F(2, 61) = 3.48$, $p < .05$, $\eta^2_p = .10$. Pairwise comparisons showed that verbal control participants produced fewer gestures than speakers with severe or moderate aphasia. There also was a main effect of Task on the total number of representational co-speech gestures, $F(1, 61) = 7.02$, $p < .01$, $\eta^2_p = .10$. More gestures were produced on the accident task. There was no significant interaction between Group and Task, $F < 1$, *n.s*. When normalizing the total number of representational co-speech gestures with the duration of the clip in seconds, there

was no significant difference in the gesture rate of severe aphasic speakers (M = .19, SD = .17), moderate aphasic speakers (M = .15, SD = .13) and verbal control participants (M = .17, SD = .15), $F < 1$, *n.s*. Neither did we find a significant difference in the gesture rate between the sweater task (M = .15, SD = .16) and the accident task (M = .19, SD = .13), $F(1, 60) = 1.90$, $p = .17$.

For the number of co-silence gestures, there was a main effect of Group, $F(2, 61) = 3.96$, $p < .05$, $\eta^2_p = .12$. There was no effect of Task, $F < 1$, *n.s.,* and no significant interaction between the two factors, $F < 1$, *n.s*. Pairwise comparisons showed that severe aphasic speakers produced more gestures without speech than moderate aphasic speakers and verbal control participants. The latter two groups did not differ significantly in the total number of co-silence gestures produced. The observed pattern was similar for the rate of co-silence gestures per second, except that the differences showed a trend toward significance, rather than reaching significance (p-values < .10).

Next we look at the relative number of gestures produced of each type. We will first present data on co-speech gestures by aphasic speakers and verbal control participants and pantomimes by nonverbal control participants. Table 4 provides an overview of the proportion of gestures produced by these groups with each technique, on either task.

Outlining/molding gestures tended to be produced more on the sweater task (M = .26, SD = .32) than on the accident task (M = .15, SD = .29), $F(1, 62) = 3.39$, p = .07, $\eta^2_p = .05$. There was a main effect of Group for the proportion of

Table 3: Means and standard deviations of the number of gestures produced with and without speech, by each group on either task (Sw. = sweater, Acc. = accident).

	Mean number of gestures per Group and Task							
	Severe aphasia		Moderate aphasia		Verbal control		Nonverbal control	
	Sw. N=12	Acc. N=12	Sw. N=13	Acc. N=13	Sw. N=8	Acc. N=9	Sw. N=9	Acc. N=8
Co-speech	2.75 (2.70)	5.25 (4.29)	2.38 (3.23)	4.92 (3.80)	.88 (1.13)	2.11 (1.76)	-	-
Co-silence	.92 (1.17)	.67 (.65)	.15 (.56)	.38 (1.39)	.00 (.00)	.11 (.33)	4.89 (1.45)	11.13 (8.87)

outlining/molding gestures, $F(3, 62) = 2.87$, $p < .05$, $\eta^2_p = .12$ (see Table 4). We did not find a significant interaction between Group and Task, $F(3, 62) = 1.60$, $p = .20$. Post hoc analyses showed that there were no significant differences between the groups on the sweater task in the proportion of outlining/molding gestures used, $F < 1$, *n.s*. On the accident task, there was a main effect of Group, $F(3, 36) = 4.64$, $p < .01$, $\eta^2_p = .28$. Pairwise comparisons showed that on the accident task speakers with severe aphasia produced a larger proportion of outlining/molding gestures than any other group.

Handling gestures were produced more on the sweater task ($M = .18$, $SD = .30$) than on the accident task ($M = .03$, $SD = .08$), $F(1, 62) = 9.69$, $p < .01$, $\eta^2_p = .14$. The main effect of Group was not significant, $F(3, 62) = 1.34$, $p = .27$ (see Table 4), and we did not find a significant interaction between Group and Task, $F < 1$, *n.s*. Post hoc analysis showed that on the accident task, there was a main effect of Group, $F(3, 36) = 5.09$, $p < .01$, $\eta^2_p = .30$. The nonverbal control participants were the only participants who made considerable use of these gestures on the accident task, significantly more so than any other group, as revealed by pairwise

Table 4: Means and standard deviations of the proportion of co-speech gestures (aphasic speakers and verbal control participants) and pantomimes (nonverbal control participants) with each representation technique, by each group on either task (Sw. = Sweater, Acc. = Accident).

	Mean proportion of gestures per group and task							
	Severe aphasia		Moderate aphasia		Verbal control		Nonverbal control	
	Sw. N=12	Acc. N=12	Sw. N=13	Acc. N=13	Sw. N=8	Acc. N=9	Sw. N=9	Acc. N=8
Outlining/Molding	.32 (.37)	.38 (.44)	.11 (.20)	.09 (.15)	.33 (.47)	.00 (.00)	.29 (.27)	.06 (.12)
Handling	.12 (.24)	.04 (.07)	.13 (.35)	.00 (.00)	.33 (.47)	.00 (.00)	.23 (.22)	.12 (.15)
Object/Enacting	.00 (.00)	.07 (.18)	.00 (.00)	.06 (.12)	.00 (.00)	.00 (.00)	.05 (.10)	.48 (.15)
Deictic	.56 (.41)	.51 (.41)	.77 (.37)	.85 (.29)	.33 (.47)	1.00 (.00)	.43 (.17)	.34 (.16)

comparisons. For the sweater task, there were no significant differences in the proportion of handling gestures between the groups, $F < 1$, *n.s*.

Object/enacting gestures were produced more on the accident task ($M = .13$, $SD = .22$) than on the sweater task ($M = .02$, $SD = .06$), $F(1, 62) = 26.42$, $p < .001$, $\eta^2_p = .30$. There also was a main effect of Group, $F(3, 62) = 21.34$, $p < .001$, $\eta^2_p = .51$ (see Table 4), and a significant interaction between Group and Task, $F(3, 62) = 13.44$, $p < .001$, $\eta^2_p = .39$. On the accident task, there was a main effect of Group, $F(3, 36) = 23.11$, $p < .001$, $\eta^2_p = .66$. Pairwise comparisons showed that on this task, nonverbal control participants produced a larger proportion of *object/enacting* gestures than any other group. Post hoc analyses showed that on the sweater task, the main effect of group was not significant. In the pairwise comparisons, the differences between nonverbal control participants and either group of aphasic speakers showed a trend toward significance.

Deictic gestures were produced more on the accident task ($M = .68$, $SD = .36$) than on the sweater task ($M = .55$, $SD = .37$), $F(1, 62) = 4.02$, $p < .05$, $\eta^2_p = .06$. There also was a main effect of Group, $F(3, 62) = 6.45$, $p < .001$, $\eta^2_p = .24$ (see Table 4), and the interaction between Group and Task was significant, $F(3, 62) = 4.07$, $p < .01$, $\eta^2_p = .17$. Post hoc analyses showed that on the accident task, there was a main effect of Group, $F(3, 36) = 11.65$, $p < .001$, $\eta^2_p = .49$. Pairwise comparisons showed that moderate aphasic speakers and verbal control participants produced a larger proportion of deictic gestures than severe aphasic speakers and nonverbal control participants. On the sweater task, the main effect of Group was not significant, $F(3, 26) = 1.90$, $p = .16$. However, pairwise comparisons showed that moderate aphasic speakers tended to produce more deictics than verbal control participants ($p = .05$) and nonverbal control participants ($p = .06$).

Since the number of co-silence gestures produced was very small for the aphasic speakers and the verbal control participants, it would not be sensible to carry out a similar analysis on these gestures. Instead, Table 5 shows the total number of co-silence gestures produced with each representation technique, by each group on either task. Some of these numbers are based on only very few speakers. We describe some special cases in the following section.

Table 5: Total number of co-silence gestures produced with each representation technique, by each group on either task (Sw. = Sweater, Acc. = Accident).

	Number of co-silence gestures per group and task							
	Severe aphasia		Moderate aphasia		Verbal control		Nonverbal control	
	Sw. N=12	Acc. N=12	Sw. N=13	Acc. N=13	Sw. N=8	Acc. N=9	Sw. N=9	Acc. N=8
Outlining/Molding	6	5	2	0	0	0	13	6
Handling	1	0	0	0	0	0	10	17
Object/Enacting	0	1	0	5	0	0	2	41
Deictic	4	2	0	0	0	1	19	25
Total	11	8	2	5	0	1	44	89

Discussion

Although the patterns of what representation techniques were used in gesture look somewhat similar across the four groups on the verbally easier sweater task, this was clearly not the case on the more difficult accident task. Apart from the verbal difficulty level of the tasks, there are other differences between the tasks that may account for this difference. On the sweater task, nonverbal control participants made frequent use of outlining/molding and handling to replace speech, whereas on the accident task, they relied far more on object/enacting. Also, on the sweater task, there was not much difference in the techniques used by verbal and nonverbal control participants, whereas on the accident task there was. The verbal control participants only produced deictic gestures on the accident task. While the moderate aphasic speakers mostly used deictic gestures on this task as well, the severe aphasic speakers produced relatively many outlining/molding gestures. This may indicate that the severe aphasic speakers were trying to express information in gesture in ways different from the verbal control participants, possibly because they needed to rely more on gesture (Tradeoff Hypothesis). Yet it seems that most severe aphasic speakers could not use the technique of object/enacting to do so, which nonverbal control participants preferred for replacing speech on the accident task. Thus, it may be

the case that the severe aphasic speakers could use outlining/molding gestures still, but had difficulty in producing object/enacting gestures. This may also hold for handling gestures. Although the severe aphasic speakers used some handling gestures on the sweater task, they made relatively little use of this technique on the accident task. In the absence of speech, severe aphasic speakers also produced many more outlining/molding gestures than object/enacting and handling gestures.

The nonverbal control participants made frequent use of object/enacting and handling gestures to describe the car accident. For example, they used their hand to represent a biker that first drove and then fell (the hand changing orientation), or they illustrated the collision between the car and the biker by letting their hands collide. Aphasic speakers did not tend to use these object/enacting techniques, with one notable exception, which we describe below. The nonverbal control participants also held their hands as though steering a car or a bike (handling). In our data sample, aphasic speakers never did this. So the aphasic speakers did not make much use of the techniques of object/enacting and handling to depict vehicles and relevant actions, despite these techniques being very suitable to do so, and despite their difficulty/ inability to convey this information verbally. Interestingly, the aphasic speakers did make some use of handling on the sweater task. Possibly, the physical presence of the referent, the sweater, sometimes helped in producing a handling gesture. Yet not all handling gestures could easily be interpreted. Some were no more than a quick grabbing motion in the air.

Both the verbal and nonverbal control participants produced a considerable proportion of outlining/molding gestures on the sweater task, indicating that on this task, this technique is suitable for producing co-speech gestures as well as to replace speech. Many people were outlining features of a sweater, such as a V-neck or sleeve length, with respect to their own body. Both groups of aphasic speakers also used this technique on the sweater task, showing similarity with the control participants in the representation techniques used. Speakers with severe aphasia also used this technique in gestures that were not accompanied by speech, which they expectedly produced more than the other speaking groups.

However, neither control group used outlining/molding much on the accident task. Nonverbal control participants hardly used molding gestures to depict vehicles like cars, or bikes. Yet aphasic speakers did sometimes do this, instead of using techniques like object/enacting or handling, like the nonverbal control

participants did. Especially the severe aphasic speakers used outlining/molding relatively frequently on the accident task, whereas moderate aphasic speakers seemed to rely more on deictic gestures. This may indicate that these techniques were the only way of gestural depiction that were readily available to most aphasic speakers for depicting the information needed in the accident task.

It thus may be the case that most aphasic speakers were unable to use the techniques of handling and object/enacting to depict on the accident task. However, the verbal control participants did not use these techniques on the accident task either. Therefore, given the task, these techniques may be more common for gestures replacing speech (pantomimes) than for co-speech gestures. It would be interesting to test whether these techniques are more readily available to aphasic speakers if they are asked to fully rely on pantomime in conveying the information. Although many aphasic speakers were unsuccessful in explaining the accident scenario verbally, their attempts at speaking may have caused them to produce co-speech gestures rather than pantomimes.

There were a few cases in our data where aphasic speakers produced readily interpretable gestures using object/enacting or handling, which they first performed without speech (possibly as a pantomime) and then repeated with some speech accompanying the gesture. For example, one person with Global aphasia first made movements as though putting on a sweater without speech, and then repeated them saying 'ja, zo' ('yes, like this'). Another person with Broca's aphasia who clearly had a problem saying a word, as she later tried to use a speech computer to convey it, used the gesture where the hand represents a biker that falls, as indicated by the hand palm changing from a vertical to a horizontal orientation. This gesture too was performed without speech (initial five times), before it was performed with speech (final time). Her eye-gazing behavior added to the impression that this gesture was intended as a pantomime, as she repeatedly looked at the gesture, then to the addressee, and then back at her gesture. This may have been in order to guide the addressee's attention to her gesture (Gullberg & Kita, 2009). After using her speech computer, this speaker said 'dood' ('dead') with the gesture, but then immediately corrected herself verbally, saying 'nee' ('no'). These two cases suggest that indeed the techniques that most aphasic speakers did not use in co-speech gesturing, may be available to some still, especially when they are giving up on conveying the message

verbally, thereby using their ability to produce pantomimes, rather than co-speech gestures.

General discussion

Our perception studies showed that gestures produced by speakers with aphasia were less informative than gestures by non-aphasic speakers and by non-aphasics who used gesture instead of speech. Moreover, gestures by people with more severely impaired speech were less informative than those by people with a more moderate speech impairment. It therefore seems that aphasic speakers could not fully compensate for their impaired expressivity in speech by gesturing. This supports the theory that gesture and speech are two sides of a coin (McNeill, 2005). That being said, being able to see the speakers too, rather than just hearing them, was sometimes helpful for judging the topic of the speakers' communication. This shows that gesture contained some information that was not contained in speech, which suggests that some speakers could compensate for their speech impairment somewhat by using gesture.

We found few significant differences in the representation techniques used by aphasic and non-aphasic speakers on the sweater task, consistent with the study by Carlomagno and Cristilli (2006). However, our analysis of gesture form did show that on the accident task, speakers with severe aphasia made relatively more use of outlining/molding than speakers with moderate aphasia and also more than non-aphasic control participants who either were or were not allowed to speak. We also found that aphasic speakers used outlining/molding gestures for referents for which control participants preferred other techniques, such as for vehicles. The control participants depicted vehicles by pretending to handle them or by having their hand represent the vehicle and depicting its movement with their hand movement, whereas the aphasic speakers tended to use outlining/molding instead. Therefore, people with aphasia may not be able to use all possible techniques for depicting in gesture freely.

It seems that especially techniques which require access to conceptual knowledge of the thing depicted (handling and object/enacting), were used relatively little by people with aphasia, while techniques using perceptual features (outlining/molding) were still available. This could be explained as a problem in translating conceptual knowledge into uttered speech and gesture (see

McNeill & Duncan, 2010). However, we need to be cautious with this interpretation, since the fact that most aphasic speakers did not use these techniques in their co-speech gesturing in our data sample does not necessarily mean they are unable to use these techniques at all.

It may be the case that depicting techniques requiring the use of conceptual knowledge are still available to some aphasic speakers when using pantomime rather than co-speech gesturing. Some special cases in our data, which we described in the previous section, support this hypothesis. Previous research already found evidence that co-speech gesturing and pantomime production result from different processes (Bartolo, et al., 2003; Rose & Douglas, 2003). Our findings confirm this in a way, because none of the people in the nonverbal control condition had difficulty gesturing without speaking, whereas speaking without gesturing tends to be much harder (e.g. Hoetjes, Krahmer, & Swerts, 2009). This may be because pantomime production is not linked to speech as tightly as co-speech gesturing, which may also be why some aphasic speakers tended to use more (conceptual) techniques when producing co-silence gestures compared to co-speech gestures. We intend to test this hypothesis in a follow-up study.

The finding that people with severe aphasia use gestures that outline shapes frequently is consistent with the case study by Cocks et al. (2011), who found that LT produced this type of gesturing frequently with difficulties in speech. This finding could be of use in clinical settings. For example, such gestures may be particularly suitable for training purposes since most aphasic speakers can still produce them. Also, it may facilitate understanding when others are aware that aphasic speakers use these gestures more widely than non-aphasic speakers. A question raised by our study is whether it would improve some aphasic speakers' ability to communicate if they are taught to replace speech with gestures using techniques that are commonly used in pantomimes, such as handling an object, or representing the object with the hand. These techniques might be more readily available to some people with aphasia if they stop their attempts at verbal communication and try to use pantomime instead.

Our studies into the informativeness of gesture, and our analysis of gestural representation techniques both suggest that like speech, co-speech gesture is impaired in most people with aphasia. It therefore seems that gesture and speech production are likely to break down together. This makes it likely, though not necessary, that the processes of speech and gesture production draw on many of

the same resources, and share an underlying process (McNeill, 2005). Although further research is needed to study the links between gesture and speech production, our study contributes to the accumulating evidence that these links are tight, rather than gesture and speech production largely being separate processes. This unfortunately limits aphasic speakers' ability to communicate by means of co-speech gestures. It seems that aphasic speakers were trying to use gesture communicatively, and did so with some success on the accident task. However, their gestures were not as informative as those of non-aphasic speakers and those of non-aphasic people who were replacing speech with gesture. This may be because the aphasic speakers could not make use of all gestural representation techniques that people without aphasia employed. Yet despite these limitations, some of the gestures they produced were informative, and added information on top of speech.

Acknowledgements

We gratefully acknowledge all speakers for allowing us to analyze their data, Renske Hoedemaker for collecting the data of our control group, and Hans Westerbeek, Hanneke Schoormans, and Manon Yassa for their help in the perception studies. We thank Vera Nijveld for doing the reliability coding.

References

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language, 44*, 169-188.

Bartolo, A., Cubelli, R., Della Sala, S., & Drie, S. (2003). Pantomimes are special gestures which rely on working memory. *Brain and Cognition, 53*, 483-494.

Beattie, G., & Shovelton, H. (1999). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of Language and Social Psychology, 18*, 438-462.

Blomert, L., Kean, M. L., Koster, C., & Schokker, J. (1994). Amsterdam-Nijmegen Everyday Language Test: construction, reliability and validity. *Aphasiology, 8*, 381-407.

Butterworth, B., & Hadar, U. (1989). Gesture, speech and computational stages. *Psychological Review, 96*, 168-174.

Carlomagno, S., & Cristilli, C. (2006). Semantic attributes of iconic gestures in fluent and non-fluent aphasic adults. *Brain and Language, 99*, 104-105.

Carlomagno, S., Pandolfi, M., Marini, A., Di Iasi, G., & Cristilli, C. (2005). Coverbal gestures in Alzheimer's type dementia. *Cortex, 41*(4), 535-546.

Chui, K. (2005). Temporal patterning of speech and iconic gestures in conversational discourse. *Journal of Pragmatics, 37*(6), 871-887.

Cienki, A., & Müller, C. (2008). Metaphor, gesture, and thought. In R. W. Gibbs (Ed.), *The Cambridge Handbook of Metaphor and Thought* (pp. 483-501). Cambridge: Cambridge University Press.

Cocks, N., Dipper, L., Middleton, R., & Morgan, G. (2011). What can iconic gestures tell us about the language system? A case of conduction aphasia. *International Journal of Language & Communication Disorders, 46*(4), 423-436.

De Ruiter, J. P., Bangerter, A., & Dings, P. (In Press). The interplay between gesture and speech in the production of referring expressions: Investigating the tradeoff hypothesis. *Topics in Cognitive Science*

Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent

events nonverbally. *Proceedings of the National Academy of Sciences of the USA, 105*(27), 9163-9168.

Goodwin, C. (Ed.). (2002). *Conversation and Brain Damage*. Oxford: Oxford University Press.

Gullberg, M., & Kita, S. (2009). Attention to speech-accompanying gestures: eye movements and information uptake. *Journal of Nonverbal Behavior, 33*(4), 251-277.

Hoetjes, M., Krahmer, E., & Swerts, M. (2009). Untying the knot between gestures and speech. In B.-J. Theobald & R. Harvey (Eds.), *Proceedings of the 8th international conference on auditory-visual speech processing* (pp. 96-101). Norwich, UK: School of Computing Sciences, University of East Anglia.

Kendon, A. (1988). How gestures can become like words. In F. Potyatos (Ed.), *Crosscultural perspectives in nonverbal communication* (pp. 131-141). Toronto, Canada: Hogrefe.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language, 47*, 16-32.

Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science, 7*, 54-60.

Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical acces: A process model. In D. McNeill (Ed.), *Language and gesture* (pp. 261-283). New York: Cambridge University Press.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

McNeill, D. (2005). *Gesture and Thought*. Chicago and London: University of Chicago Press.

McNeill, D., & Duncan, S. (2010). Gesture and growth points in language disorders. In J. Guendouzi, F. Loncke & M. J. Williams (Eds.), *The handbook of psycholinguistic and cognitive processes* (pp. 663-685). New York, London: Psychology Press.

Melinger, A., & Kita, S. (2007). Conceptualisation load triggers gesture production. *Language and Cognitive Processes, 22*(4), 473-500.

Melinger, A., & Levelt, W. J. M. (2004). Gesture and the communicative intention of the speaker. *Gesture, 4*(2), 119-141.

Rose, M. L. (2006). The utility of arm and hand gestures in the treatment of aphasia. *Advances in Speech-Language Pathology, 8*(2), 92-109.

Rose, M. L., & Douglas, J. (2003). Limb apraxia, pantomime, and lexical gesture in aphasic speakers: Preliminary findings. *Aphasiology, 17*(5), 453-464.

So, W. C., Kita, S., & Goldin-Meadow, S. (2009). Using the hands to indentify who does what to whom: Gesture and speech go hand-in-hand. *Cognitive Science, 33*, 115-125.

Van der Meulen, I., Van de Sandt-Koenderman, W. M. E., Duivenvoorden, H. J., & Ribbers, G. M. (2009). Measuring verbal and non-verbal communication in aphasia: reliability, validity, and sensitivity to change of the Scenario Test. *International Journal of Language & Communication Disorders*, 1-12.

Chapter 6

General discussion and conclusion

General discussion

We set out to address the question of whether gesture is part of a speaker's attempt at communication, like speech is, or whether its role is limited to supporting a speaker's cognitive processes underlying speech production. We aimed to test proposed theories empirically, thereby contributing to their grounding in empirical data and ultimately their further development. Each of our studies shows one way in which gesture production resembles speech production.

Study 1 and 2

Our first two studies show that gesture is influenced by the knowledge a speaker has about the addressee. Even when the identity of the addressee was provided only in the instruction prior to a communication task, while during the task the conditions were exactly the same across conditions, speakers gestured less frequently toward a presumed artificial audiovisual summarizer than toward a presumed human addressee, who would summarize their narration. Similarly, independent of whether speakers could see their addressee, they gestured more when they were told that their addressee could see them. Both these results show that speakers adapt their gesturing to their beliefs about their addressee, which supports the theory that gestures are part of a speaker's communicative effort (Kendon, 2004).

Importantly, we also assessed various variables partaking to speech in these studies. After all, if speakers speak differently toward different addressees, then the speech production process may place different demands on gesturing and gesture may change as a result. In our first study, participants spoke somewhat slower to the artificial than to the human addressee. It could be that speakers experienced less time pressure when addressing an artificial addressee, resulting in them needing to produce fewer gestures for speech facilitation. In this case still, there would be an effect of the speaker's belief about the addressee on gesture production, but it may be mediated by speech production. Although we did not find any evidence for such mediation, it would be interesting to investigate the relation between time pressure and gesture rate in future work, and examine whether such mediation takes place. The mediation may also happen the other way around. That is, the beliefs a speaker holds may influence gesture production, which in turn could affect speech production. Alternatively,

the effects found in gesture and speech may both result from the speaker's beliefs about the addressee more directly, for example, because both are informed by the speaker's communicative intention, which is influenced by the speaker's beliefs. Based on our data, we think the latter explanation most likely, as explained in Chapter 2.

In the study in which we manipulated visibility by using computer mediated communication, we did not find any differences in the number of words used, the diversity of words used, or the number of filled pauses between a mediated condition in which speakers could be seen by the addressee and a mediated condition in which they could not. Yet the difference in the number of gestures produced was striking. In the analyses of this study, we used speech rate as a covariate, correcting for any effects of the speech rate on speakers' gesture rate. Still, significantly more gestures were produced when speakers knew they could be seen. This confirms that the beliefs the speaker held about the communicative setting also influenced gesture directly, rather than this effect being fully mediated by speech. Therefore, our first two studies provide evidence that gesture is also adapted to the addressee, rather than just to speech production. This suggests that gesture is part of a speaker's attempt at communication itself, instead of just being facilitative to speech. Additionally, in our perception studies related to these first two production studies, raters showed great sensitivity to the differences in gesturing of speakers in different communicative settings, as one would expect if gesture serves a communicative purpose.

Study 3

In our third study, we compared the repetition of meaningful gestures across interlocutors to the repetition of meaningful units in speech, such as words or referring expressions. First, we found that speakers only repeated *meaningful* gestures that they had observed. That is, gestures that matched the meaning of the concurrent speech during perception and production. This shows that the property of carrying propositional meaning was important for gestures being copied. If speakers would copy each other's gestures solely to express liking or to express that they belonged to the same group as another speaker, there is no reason why only meaningful gestures would be repeated. In this sense, gestures are like words or referring expressions, which are also repeated across interlocutors only if their meaning fits the current context closely enough (see for example, Brennan & Clark, 1996; Van Der Wege, 2009).

We found another similarity between speech and gesture when it comes to adaptation. Similar to a theory proposed for the repetition of referring expressions across interlocutors (Brennan & Clark, 1996), the repetition of gestures across interlocutors seems to be mediated by concepts. Speakers did not copy just any feature of a gesture they had observed. Rather, perceiving gestures influenced the way speakers conceptualized their task. Only those features of a gesture that matched a certain concept were repeated when the speaker subsequently expressed this concept. Features that did not match the concept were not repeated, but rather changed to match it. This shows that the convergence of concepts across interlocutors may underlie the convergence of gesture forms, as Brennan and Clark, among others, proposed for lexical forms.

In this third study, speech cannot have been the discriminating factor when it came to concept formation, since participants heard the same speech in each condition. Only the gestures they perceived differed. Therefore, gesture is the most probable cause for participants forming different conceptualizations of the task at hand. Thus, this study also shows that concepts can be communicated through gesture.

It is important to replicate our findings on the repetition of gesture forms across interlocutors with different paradigms in future work, to test if our results generalize to different contexts. Importantly, we do not mean to imply that no aspect of gesture is ever copied across interlocutors without a representation of meaning driving this process. We only looked at gestures that were meaningful in an iconic way. Different processes may underlie the copying of other types of gestures and of other aspects of gesture, such as gesture rate, and how information is structured in gesture. Our study is one of the first studies to systematically look at adaptation in gesture. Much more work is needed in this area. Yet the gestures we examined seemed to behave just like words when it came to the adaptation of one interlocutor to another.

Study 4

Our final study assessed the similarity between gesture and speech when it comes to the processes underlying their production. We studied gestures produced by aphasic speakers, who had severe or moderate difficulty expressing themselves verbally, as a result of brain damage.

In a perception study, we found that gestures of speakers with severe aphasia were less informative than those of speakers with moderate aphasia, which in

turn were less informative than those of healthy controls. This shows that rather than gesture taking on more of the communicative burden when speech is impaired, it tended to degrade with speech.

Next, we looked at gesture form more closely. We compared gestures produced by aphasic speakers to those of healthy controls, who either were or were not allowed to speak. The aphasics were using more ways of representing in gesture than the controls who could speak, indicating that they may have tried to communicate more information through gesture. Yet we found that aphasic speakers did not use the same ways of representing in gesture as the controls who were asked to communicate through gesture instead of speech. The aphasics hardly depicted referents by pretending to handle them, or by having their hands represent the referent. Instead, they produced relatively many gestures that outlined shapes. This may indicate that they had difficulty with gestures that express conceptual rather than perceptual information.

This may be because especially gestures expressing conceptual information co-develop with speech from an idea unit into a bimodal utterance, the way McNeill describes the co-production of gesture and speech in his growth-point theory (McNeill, 2005). In the model by Kita and Özyürek (2003), gesture is not only informed by the conceptualization stage of language production, but also more directly by the spatio-motor store in working memory. Gestures expressing perceptual features may rely on the content of working memory more heavily than gestures expressing conceptual information. In future work, we intend to test this hypothesis and to examine whether the gestures that healthy speakers produce during word finding problems resemble the gestures produced by aphasic speakers in that they depict perceptual rather than conceptual information.

Both our perception study and our analyses of gesture forms provide evidence that gesture degrades with speech in aphasia. That is, gesture production tended to be impaired when speech production was. This does not show conclusively that gesture and speech production partly rely on the same resources, or that gesture and speech are two outcomes of a single process. They could still be separate processes that rely on different parts of the brain, which happened to be damaged at the same time in most of the patients whose gestures we examined. It does seem though that gesture and speech production are likely to break down together, which makes it more likely that the underlying production processes are closely linked.

In future work, it would be interesting to test whether aphasic speakers can use more representation techniques still and perhaps gesture more informatively when they rely on their ability to produce pantomimes rather than co-speech gestures. Pantomimes are gestures that are produced without speech and that are intended to be intelligible without speech. Since pantomime production is likely to be linked to speech less tightly than co-speech gesturing, some aphasic speakers may be more successful when expressing themselves through pantomime than through co-speech gesturing. It would also be interesting to see if there are differences in the extent to which gesture is impaired between speakers with verbal problems on different linguistic levels, such as the phonological and the semantic level. This may give insight into the question at what levels gesture and speech production are linked.

Conclusion

In sum, the four studies in this dissertation show various ways in which speech and co-speech gesture production are alike. Both speech and gesture can be produced with a communicative intent, both bear a close link to meanings and concepts, and both may break down together in aphasia. The fact that gestures can be part of speakers' communicative effort does not limit the functional roles of gesture to communication. Yet based on our findings, we argue that it is not limited to serving speech or serving a speaker's cognition either. Rather, we consider gesture to be part of language itself. Gesture is language in the hands.

References:

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology-Learning Memory and Cognition, 22*(6), 1482-1493.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language, 47*, 16-32.

McNeill, D. (2005). *Gesture and Thought*. Chicago and London: University of Chicago Press.

Van Der Wege, M. M. (2009). Lexical entrainment and lexical differentiation in reference phrase choice. *Journal of Memory and Language, 60*, 448-463.

Summary

This dissertation addresses the question of whether gesture is part of a speaker's attempt at communication, like speech is, or whether it rather serves speaker-internal purposes, like facilitating speech production and other cognitive processes. Each of our four empirical studies shows one way in which gesture production resembles speech production.

Study 1

In our first study, we tested whether speakers produce equally many and similar gestures when addressing an artificial addressee, as when addressing a human addressee. The novelty of this approach is that we only varied the beliefs a speaker had about the addressee, while the speaker's environment and other characteristics of the communicative setting were exactly the same across conditions. Speakers were asked to retell the story of an animated cartoon they had watched, while being seated in front of a camera. Beforehand, they were either told that the camera recording was shown to a person in another room, or that it was used as input to an audiovisual speech recognizer that was located in another room.

Our reasoning was that if speakers produce some gestures for their addressee, then their beliefs about the addressee are likely to influence their gesture production, whereas if gesturing solely serves speech production, such a difference would not be expected. We found that whether speakers thought to be addressing another person or an audiovisual speech recognition system influenced their gesture production. Speakers produced fewer gestures per word and fewer large gestures toward the presumed artificial system. In a subsequent perception study, we found that this difference in gesture production could be interpreted meaningfully by other participants, who reliably judged whether a speaker had been addressing a human or an artificial addressee, from seeing the speaker's gesturing alone. These results support the hypothesis that speakers gesture partly for their addressee, that is, with a communicative intent.

Study 2

In our second study, we further explored what knowledge about their addressee speakers apply to their gesturing. Previously, it had been found that speakers gesture less frequently when interlocutors cannot see each other. However, it was unknown whether this resulted from speakers using their knowledge that the addressee could not see them, or from speakers not seeing their addressee.

By means of computer-mediated communication, we created four settings in which each interlocutor either could or could not see the other interlocutor. This way, we independently assessed the effects of seeing the addressee and being seen by the addressee on gesture production. We used the same cartoon narration task as in our first study.

We found that speakers gestured more when they knew their addressee could see them, independent of whether they could see their addressee. This shows that speakers applied their knowledge of their addressee's visual perspective to their gesturing. While being seen always increased speakers' gesture rate, seeing the addressee only increased gesture production if the mediated setting allowed for natural gazing behavior. An additional perception study showed that an increased gesture rate was associated with higher expressivity. In sum, knowing that their addressee can see them causes speakers to produce more gestures and thereby to be more expressive than when they know they cannot be seen. This is additional support for the hypothesis that some gestures are intended communicatively.

Study 3

In our first two studies we found that gesture resembles speech in that gesture can be intended communicatively. In our third study, we examined how a phenomenon that is well known in speech, the adaptation of one interlocutor's communicative behavior to another's, occurs in gesture. Studying adaptation in gesture not only tells us something about the extent to which gesture and speech are similar, but it can also inform us on what mechanisms underlie adaptation.

First, we tested whether the repetition of representational gestures across interlocutors is related to these gestures carrying propositional meaning. Participants saw video clips in which a speaker performed a gesture during a narration. Half the participants saw the speaker perform gestures that matched the content of his concurrent speech. For example, he moved his hands as though running while talking about someone running away. The other half of the participants saw the speaker perform gestures that did not match the content of his speech. For example, he performed the running gesture while he was talking about someone looking through binoculars. We found that only those gestures that matched the content of concurrent speech were likely to be repeated by participants in their subsequent retellings of the narrations. This shows that a gesture's meaning plays a critical role in its repetition across interlocutors.

We subsequently manipulated the correspondence between a gesture's form and meaning more subtly. A confederate and a participant took turns describing routes that were presented to them on little maps. The confederate either gestured as though pointing out the route on a vertically oriented map, or as though following the route through a (horizontal) city. In a pre-study, we found that when pointing on a map, people tend to point with one finger extended, while when pointing out a route in the streets both this hand shape and a hand shape with all fingers extended are used. Therefore, the confederate independently varied whether she gestured with all or just one finger extended as an index. We measured whether participants copied the confederate's hand shape and perspective (vertical map or horizontal route) in their own gesturing.

We found that participants adapted to the confederate's perspective. When the confederate gestured in the vertical map perspective, participants were more likely to do so as well. However, participants adapted to the confederate's hand shape only if she gestured in the horizontal route perspective. This can be explained in terms of meaning. In the horizontal route perspective, both hand shapes are commonly used in the Netherlands for giving directions. Therefore, participants could and did adapt to meaningful gestures. However, when pointing on a map, it is far more common to point with one finger extended as opposed to four. Therefore, if the confederate's vertical gestures led participants to think of the task as describing the route on a map, we would expect them to gesture with one finger, independent of the confederate's hand shape. This is indeed what we found. On the other hand, if participants would copy the confederate's movements without ascribing meaning to them, this would not be expected. Therefore, these results further support the hypothesis that it is not just a gesture's form that is being copied, but also its meaning.

In sum, we found that only meaningful gestures, and only meaningful aspects of gestures were copied across interlocutors. This suggest that the repetition of gesture forms across interlocutors is a result of concepts converging across interlocutors, rather than it being automated copying of form. This has also been suggested for the repetition of referring expressions across interlocutors. Speech and gesture thus seem to act alike when it comes to adaptation.

Study 4

We have already seen that like speech, gesture can be intended communicatively and that meanings can converge across interlocutors through gesture as well as

through speech. These similarities between gesture and speech raise the question of how closely gesture and speech production are linked to one another. In our fourth and final study, we aimed to shed light on this issue by examining gestures produced by aphasic speakers.

First, we examined the intelligibility of aphasic speakers' gestures by means of perception studies. Raters watched video-clips played without sound and were asked to judge whether the speaker was trying to communicate about buying a sweater or about a traffic accident. We found that gestures of speakers with severe aphasia were less informative than those of speakers with mild aphasia, which in turn were less informative than those of healthy control participants. This shows that rather than gesture taking on more of the communicative burden when speech was impaired, it tended to degrade with speech.

Second, we looked at the representation techniques employed in gesture. We found that aphasic speakers did not use all the techniques used by healthy control participants who were asked to communicate with gesture instead of speech. This showed especially when talking about the traffic accident and referring to vehicles. Instead of pretending to handle a vehicle or representing it with a hand like the gesturing control participants did, people with severe aphasia frequently used gestures that outlined shapes. People with severe aphasia also used outlining gestures relatively more frequently than speakers with more moderate aphasia and healthy speakers. It seemed that aphasic speakers were trying to express meaning in gesture, but could not use gesture as freely as healthy control participants.

Both our perception study and our analyses of representation techniques provide evidence that gesture degrades with speech in aphasia. This supports theories and frameworks in which the production of gesture and speech are interrelated and share common resources.

Conclusion

The four studies in this dissertation show various ways in which speech and co-speech gesture production are alike. Both speech and gesture can be produced with a communicative intent, both bear a close link to meanings and concepts, and both may break down together in aphasia. The fact that gestures can be part of speakers' communicative effort does not limit the functional roles of gesture to communication. Yet based on our findings, we argue that it is not limited to serving speech or serving a speaker's cognition either. Rather, we consider gesture to be part of language itself. Gesture is language in the hands.

Acknowledgements

An ordinary day in the lab… I turn up the lights, take two cameras from the cabinet and set them up in the room. Throughout the day, the hand gestures participants produce are captured on videotapes. Those I digitalize in order to analyze the captured movies on my computer, playing them back frame-by-frame, again and again, allowing me to study participants' movements in great detail. How different from the start of science in Europe, when one had to undertake an expensive and rather dangerous trip to Paris even to obtain a reliable ruler. Needless to say there are an endless number of scientists without whom the work presented in this dissertation would have never been accomplished by me. I admire their dedication to science and I have gratefully made use of their accomplishments.

This holds for three scholars in particular. I like to seize this opportunity to thank Emiel Krahmer, Fons Maes and Marc Swerts for supporting and inspiring me throughout my PhD project. I also thank Tilburg University for generously funding my PhD project as well as my academic development.

It was Emiel who initially suggested the topic of gesture to me. Perhaps that was his one suggestion that I did not fight. Emiel has always been available for rigorous scientific discussion, which I very much appreciate. Also, Emiel initiated many of my experiences abroad, by encouraging me to attend international courses and conferences. Moreover, he is an inspiring researcher with unbelievable time management skills, and everlasting optimism. I thank Emiel for being such a reliable and inspiring advisor.

Fons has been very supportive to me during the first few years of my project, when I was still overwhelmed by all the different opinions and ideas of my three advisors. He was first to hear me out and I felt he was most likely to side with me, supporting me to pursue my own insights. As the leader of the department I worked in, I feel he had a major part in me being able to work in a fun and cooperative environment throughout this project. Fons is a very inspiring leader, whom I felt I could turn to with anything that kept me from reaching my academic goals. As a researcher, Fons never let me get away with being fuzzy, or putting down only half of my thoughts, which greatly improved my papers. Yet I foremost thank Fons for going the extra mile in supporting me in this project.

Marc I admire mostly for his creativity. Marc can look at scientific problems from infinitely many angles. No matter how impossible a question seemed or how stuck I was, Marc could always make me see yet another way of

approaching the problem. Also, Marc has played a major role in me developing some confidence as a scientist. His endless streams of positive feedback have kept me going when the going got tough. I thank Marc for showing so much faith in me. As a researcher, Marc has inspired me with his achievements as well as his perseverance, setting an example to live up to. Marc also has excellent mediation skills, which helped this arrangement of having three full professors as thesis advisors to work out like it did. I could not have wished for more!

I gratefully thank all members of my PhD committee: Susan Brennan, Jan de Ruiter, Sotaro Kita, Asli Özyürek, and Mieke van de Sandt-Koenderman, for their careful reading of my manuscript and their insightful comments to it. I look forward to discussing my work with them at my thesis defense. Kita I also thank for always asking the right question after one of my presentations, and repeating it until I saw its import. This has greatly helped me direct my studies. I am also thankful for the opportunities I got to present at other labs, which enabled me to meet many scholars working on similar and related topics.

Over the years, I have come to know many researchers in the field of Gesture. I am thankful for their warm welcome, their enthusiasm and their constructive comments. I like to thank all reviewers and editors who helped me improve the papers that this dissertation is based on, in particular, I thank Jan de Ruiter, Adam Kendon and Victor Ferreira.

Since I have not included any acknowledgements in my master's thesis, I still like to thank Rineke Verbrugge and Petra Hendriks, who supervised me at the time. They provided a fascinating research question to me and I much enjoyed the freedom they allowed me in approaching it. As highly successful women in science, they are inspiring role models to me and they never hesitated to help me advance in my academic career. In this light, I also thank Niels Taatgen and John Anderson, for helping me obtain my first research position right after finishing my master's degree. Working and living abroad has been an invaluable experience to me.

During this PhD project, I got the opportunity to advise Nelianne van den Berg in writing her master's thesis. I greatly enjoyed this process and I was inspired by her discipline and work pace. Nelianne collected and coded part of the data in Chapter 3. I am thankful for this fruitful cooperation. I also like to thank Vera Nijveld, who repeatedly helped me out with reliability coding and I thank Hanneke Schoormans for transcribing very many videos.

I thank all participants to my studies for allowing me to analyze their videos. It happened to me quite a few times that I greeted someone a bit too enthusiastically in the hallways of Tilburg University or in the city of Tilburg itself. After analyzing their gestures for about an hour, it simply felt like I knew people a lot better than I actually did.

I gratefully thank Hans Westerbeek for designing the cover of this dissertation for me. If it were not for him this thesis would have looked completely boring. He has been incredibly patient with all my concerns about minor issues that no one else will actually notice. I also thank Hans for helping me out time and time again as our ZLAU (very local IT support).

I thank Carel van Wijk for teaching me how to apply statistics to my data and Lennard van der Laar for teaching me how to work with video files. I thank Bernd Hellema for patiently helping me with numerous technical issues and for writing the software for my line of five failed pilot studies (which I still believe will work out some day, but just in case). I thank Rein Cozijn for managing and improving the lab at Tilburg University and for getting the Eye-Catchers and Noldus Observer. Although it remains a time-consuming process, the latter sped up my gesture coding considerably. I thank the supportive staff at Tilburg University for all the ways in which they enable me to do my job.

I thank all my coworkers at Tilburg University for their help and their excellent company, especially during our Thursday cookie meetings and the walks in the woods at lunchtime. In particular I thank my roomie Lisanne van Weelden, both for distracting me from my work and for urging me on with it. I thank all my fellow PhD students for taking part in my pilots, listening to my stories, and the lively PhDinners we had. I thank Lauraine Sinay for giving our 'fourth floor' a more homely feel to it. It is just not the same when she is not there. Not in the least I thank both Marieke Hoetjes and Lisanne van Weelden for being my paranymphs and for supporting me as enthusiastically as they do.

It is a little silly to thank my friends and other special people here. I hope they all know what they mean to me without me writing a dissertation. Nevertheless, I like to thank everyone who travels or traveled with me for part of the way, including Frits Knaack, Bernd Hellema, Carina Pals, Nathalie Bastiaansen, Sandra van der Helm, Robin Waart, Chistine Kühnel, Saskia Hutten, the Ruijter family, Jef and Bettina Bakker, Jos, Maria and Martijn de Koning, Adinda Visser, Harry de Roest and Marian van Item, Jan and Marita Mol, Lennie Mol and Hans Banens, Terry Chung, Michelle Gusic, Monique van

de Sande, Maaike Warmerdam, Rob Leereveld, Annouck Leest, Loes van Benschop, Jessica Schouwenaars, Carel Veerman, Hennie van Woudenberg, John Keijzer, Viola van Wijngaarden, Janine van Trierum, Susan Zwakkenberg, Debby de Waal, Esmé Smits, Hans Kuijvenhoven, Dorinda Vredeveldt, Marjolein van Schaik – van den Berge, Henriëtte and René de Koning, Gerline and Marco Bakker and many others.

Publication list

Journal publications

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (In Press). Adaptation in gesture: Converging hands or converging minds? *Journal of Memory and Language.*

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (In Press). Seeing and Being Seen: The effects on gesture production. *Journal of Computer-Mediated Communication.*

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2009). The communicative import of gestures: Evidence from a comparative analysis of human-human and human-machine interactions. *Gesture,* 9(1), 97-126.

Verbrugge, R., & Mol, L. (2008). Learning to apply Theory of Mind. *Journal of Logic, Language and Information,* 17(4), 489-511.

Working papers

Mol, L., Krahmer, E., & Van de Sandt-Koenderman, W. M. E. (Submitted). Gesturing by aphasic speakers, how does it compare?

Papers in conference proceedings (peer reviewed)

Mol, L., Krahmer, E., & Van de Sandt-Koenderman, W. M. E. (2011). Gesturing by aphasic speakers, how does it compare? In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1454-1459). Austin, TX: Cognitive Science Society.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2010). Converging Hands or Converging Minds? In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 115-120). Austin, Tx: Cognitive Science Society.

Mol, L., Krahmer, E. (2010). Handling what the other sees: The effects of seeing and being seen on gesture production. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 115-120). Austin, Tx: Cognitive Science Society.

Mol, L., Krahmer, E., & Swerts, M. (2009). Alignment in iconic gestures: Does it make sense? In B-J. Theobald & R. Harvey (Eds.), *Proceedings of the 8th International Conference on Auditory-Visual Speech Processing (AVSP 2009)* (pp. 3-8). Norwich, UK: School of Computing Sciences.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2009). Communicative gestures and memory load. In N.A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 1569-1574). Austin, TX: Cognitive Science Society.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2008). Gesticulation and audience design. In *Proceedings of the 3rd international conference on cognitive science* (110-111). Moscow: Art and Publishing Center.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2007). In J. Vroomen, E. J. Krahmer, & M. Swerts (Eds.), *Proceedings of the 6th International Conference on Auditory-Visual Speech Processing (AVSP 2007)* (pp. 200-205). Tilburg: Tilburg University.

Mol, L., Taatgen, N., Verbrugge, R., & Hendriks, P. (2005). Reflective Cognition as a Secondary Task. In: B.G. Bara, L. Barsalou, and M. Bucciarelli (Eds), *Proceedings of the 27th Annual Meeting of the Cognitive Science Society* (pp. 1525-1530). Mahwah, NJ: Erlbaum.

Mol, L., Verbrugge, R., & Hendriks, P. (2005). Learning to reason about other people's minds. In *Proceedings of the Joint Symposium on Virtual Social Agents*, SSAISB 2005 (pp. 191-198). Hatfield, UK: The Society for the Study of Artificial Intelligence and the Simulation of Behaviour.

Abstracts of spoken papers (peer reviewed)

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2010). Converging Hands or Converging Minds? Presented at: 'Gestures, evolution, brain, and linguistic structures': the 4th international conference of the International Society for Gesture Studies (ISGS), Frankfurt (Oder), Germany.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2008). Audience Design en Handgebaren. Presented at: VIOT, Amsterdam, The Netherlands.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2008). Cognitive Effort and Gesturing. Presented at: The Workshop on Speech and Face-to-Face

Communication dedicated to the memory of Christian Benoît, Grenoble, France.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2008). Look who's being talked to. Presented at: Language, Communication, and Cognition (LCC), Brighton, United Kingdom.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2007). The communicative import of gestures: Evidence from a comparative analysis of human-human and human-machine interactions. Presented at: 'Integrating Gestures': the 3rd international conference of the International Society for Gesture Studies (ISGS), Evanston, IL, USA.

Mol, L., Taatgen, N., & Anderson, J. (2005). Individual Differences in Multi-Tasking. Presented at: The 12th Annual ACT-R Workshop, Trieste, Italy.

TiCC Ph.D. series

1. Pashiera Barkhuysen. *Audiovisual Prosody in Interaction.* Promotores: M.G.J. Swerts, E.J. Krahmer. Tilburg, October 3, 2008.
2. Ben Torben-Nielsen. *Dendritic morphology: function shapes structure.* Promotores: H.J. van den Herik, E.O. Postma. Co-promotor: K.P. Tuyls. Tilburg, December 3, 2008.
3. Hans Stol. *A framework for evidence-based policy making using IT.* Promotor: H.J. van den Herik. Tilburg, January 21, 2009.
4. Jeroen Geertzen. *Dialogue act recognition and prediction.* Promotor: H. Bunt. Co-promotor: J.M.B. Terken. Tilburg, February 11, 2009.
5. Sander Canisius. *Structured prediction for natural language processing.* Promotores: A.P.J. van den Bosch, W. Daelemans. Tilburg, February 13, 2009.
6. Fritz Reul. *New Architectures in Computer Chess.* Promotor: H.J. van den Herik. Co-promotor: J.W.H.M. Uiterwijk. Tilburg, June 17, 2009.
7. Laurens van der Maaten. *Feature Extraction from Visual Data.* Promotores: E.O. Postma, H.J. van den Herik. Co-promotor: A.G. Lange. Tilburg, June 23, 2009 (cum laude).
8. Stephan Raaijmakers. *Multinomial Language Learning.* Promotores: W. Daelemans, A.P.J. van den Bosch. Tilburg, December 1, 2009.
9. Igor Berezhnoy. *Digital Analysis of Paintings.* Promotores: E.O. Postma, H.J. van den Herik. Tilburg, December 7, 2009.
10. Toine Bogers. *Recommender Systems for Social Bookmarking.* Promotor: A.P.J. van den Bosch. Tilburg, December 8, 2009.
11. Sander Bakkes. *Rapid Adaptation of Video Game AI.* Promotor: H.J. van den Herik. Co-promotor: P. Spronck. Tilburg, March 3, 2010.
12. Maria Mos. *Complex Lexical Items.* Promotor: A.P.J. van den Bosch. Co-promotores: A. Vermeer, A. Backus. Tilburg, May 12, 2010 (in collaboration with the Department of Language and Culture Studies).

13. Marieke van Erp: *Accessing Natural History. Discoveries in data cleaning, structuring, and retrieval*. Promotor: A.P.J. van den Bosch. Tilburg, June 30, 2010.

14. Edwin Commandeur: *Implicit Causality and Implicit Consequentiality in Language Comprehension*. Promotores: L.G.M. Noordman, W. Vonk. Co-promotor: R. Cozijn. Tilburg, June 30, 2010.

15. Bart Bogaert: *Cloud Content Contention*. Promotores: H.J. van den Herik, E.O. Postma. Tilburg, March 30, 2011.

16. Xiaoyu Mao: *Airport under Control.* Promotores: H.J. van den Herik, E.O. Postma. Co-promotores: N. Roos, A. Salden. Tilburg, May 25, 2011.

17. Olga Petukhova: *Multidimensional Dialogue Modelling*. Promotor: H. Bunt. Tilburg, September 1, 2011.

18. Lisette Mol: *Language in the hands.* Promotores: F. Maes, E.J. Krahmer, M.G.J. Swerts. Tilburg, November 7, 2011.